View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by D-Scholarship@Pitt

DEVELOPMENT AND EVALUATION OF A NOVEL HUMAN-SPECIFIC FECAL SOURCE TRACKING MARKER

by

Elyse Stachler

Bachelor of Science in Civil Engineering, University of Notre Dame, 2013

Master of Science in Civil Engineering, University of Pittsburgh, 2015

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Elyse Stachler

It was defended on

December 13, 2017

and approved by

Kyle J. Bibby, Ph.D., Assistant Professor, Civil and Environmental Engineering Department

Radisav D. Vidic, Ph.D., Professor, Civil and Environmental Engineering Department

Leanne Gilbertson, Ph.D., Assistant Professor, Civil and Environmental Engineering Department

Emily Elliot, Ph.D., Associate Professor, Department of Geology and Environmental Science

Dissertation Director: Kyle J. Bibby, Ph.D., Assistant Professor, Civil and Environmental Engineering Department

Copyright © by Elyse Stachler

2018

DEVELOPMENT AND EVALUATION OF A NOVEL HUMAN-SPECIFIC FECAL SOURCE TRACKING MARKER

Elyse Stachler, Ph.D.

University of Pittsburgh, 2018

Enteric viruses pose a significant health risk to the public through exposure to contaminated environmental waters, costing society billions of dollars to treat millions of excess illnesses annually. Current available methods to monitor human fecal pollution in impacted waters either do not correlate with viral pathogen presence and risk, cross-react with other animal sources, or are not abundant enough in environmental waters to be reliably detected. This dissertation investigates a novel target for a human-specific indicator of fecal pollution, the bacteriophage "crAssphage". First, a metagenomic evaluation was conducted as an initial step to evaluate the potential of the crAssphage genome for assay development. Results indicated that crAssphage is enriched in sewage and more abundant than other viruses and bacteriophages in sewage, suggesting its high potential as a target for human-specific marker development. Next, end-point PCR primers were designed along the length of the crAssphage genome and screened against human sewage samples and non-human animal fecal samples to assess which genomic regions may be the most useful as source tracking markers. The two best performing primer pairs were adapted to TaqMan qPCR assays. These assays were evaluated head-to-head against two bacterial qPCR human source tracking markers to evaluate their performance. The assays were widespread in geographically diverse human sources and as abundant or more abundant in

sewage and impaired environmental water than the bacterial-based assays. Moreover, the assays displayed high human-association. Finally, an initial environmental evaluation was conducted to correlate the crAssphage-based assays to pollution events, culturable bacterial and phage indicators, and molecular bacterial and viral indicators. This study demonstrated the usefulness of the crAssphage assays in a real-world system, validating their high abundance and correlation with pollution events and other indicators of fecal pollution. Ultimately, the research in this dissertation contributes two novel viral-based technologies for detection of human fecal pollution that will enhance management of environmental waters and protect public health through the development of an abundant viral fecal source-tracking marker.

TABLE OF CONTENTS

1.0 INTRODUCTION ... 1

1.1 MOTIVATION ... 1

1.2 RESEARCH OBJECTIVES ... 4

2.0 A REVIEW OF HUMAN SOURCE TRACKING TECHNOLOGIES: STATE OF THE SCIENCE AND RESEARCH NEEDS ... 8

2.1 MICROBIAL SOURCE TRACKING INTRODUCTION ... 8

2.1.1 Shortcomings of the Current Approach ... 8

2.1.2 Origins of MST ... 9

2.1.3 Terms and definitions ... 10

2.2 BACTERIAL MST ASSAYS ... 11

2.2.1 *Bacteroides* 16S rRNA Assays ... 11

2.2.1.1 HF183 ... 12

2.2.1.2 BacH ... 13

2.2.1.3 HuBac ... 14

2.2.1.4 BacHum ... 14

2.2.1.5 HumanBac1 ... 15

2.2.1.6 BacHuman ... 15

2.2.2 Other *Bacteroides* Assays ... 16

2.2.3 Other Bacterial Targets ... 17

2.2.4 Challenges and Limitations 17

2.3 VIRAL MST ASSAYS 18

2.3.1 Enteric Pathogen Viral Assays 19

2.3.1.1 Norovirus 19

2.3.1.2 Adenovirus 20

2.3.1.3 Enterovirus 21

2.3.2 Non-Pathogenic Viral Targets 21

2.3.2.1 Human Polyomavirus 22

2.3.2.2 *Pepper Mild Mottle Virus* 23

2.3.3 Challenges and Limitations 24

2.4 BACTERIOPHAGE MST ASSAYS 25

2.4.1 F^+ RNA Coliphage 26

2.4.2 Phages of *Bacteroides* 27

2.4.3 Challenges and Limitations 28

2.5 RESEARCH NEEDS 28

2.5.1 Ideal Marker Characteristics 29

3.0 METAGENOMIC EVALUATION OF THE HIGHLY ABUNDANT HUMAN GUT BACTERIOPHAGE CRASSPHAGE FOR SOURCE TRACKING OF HUMAN FECAL POLLUTION 31

3.1 INTRODUCTION 32

3.2 MATERIALS AND METHODS 34

3.2.1 Genome downloads 34

3.2.2 Identification of metagenomes 34

3.2.3 Alignment to the reference genomes 35

3.3 RESULTS AND DISCUSSION 36

4.0 QUANTITATIVE CRASSPHAGE PCR ASSAYS FOR HUMAN FECAL POLLUTION MEASUREMENT 43

4.1 INTRODUCTION 44

4.2 MATERIALS AND METHODS 47

4.2.1 Sample Collection 47

4.2.2 DNA Extraction and Quantification 48

4.2.3 Selection of Candidate Genetic Regions for PCR-Based Assay Development 49

4.2.4 Candidate Primer Set Design 50

4.2.5 End-point PCR Amplifications 50

4.2.6 Candidate End-Point PCR Primer Set Evaluation 51

4.2.7 DNA Sequence Verification of Top Performing End-Point PCR Primer Sets 53

4.2.8 CrAssphage qPCR Assay Development 53

4.2.9 qPCR Amplifications 54

4.2.10 qPCR Standard DNA Material Preparation 54

4.2.11 Performance Testing of crAssphage qPCR Assays 55

4.2.12 Data Analysis 56

4.3 RESULTS 57

4.3.1 Putative Human-Associated crAssphage Genetic Regions and Candidate Primer Set Design 57

4.3.2 Identification of Human-Associated crAssphage Genetic Regions with End-Point PCR 59

4.3.3 DNA Sequencing Verification 60

4.3.4 Performance of crAssphage qPCR Assays 61

4.3.5 Experiment Controls 64

4.4 DISCUSSION 64

4.4.1 Identification of Human-Associated CrAssphage Genetic Regions 64

4.4.2 Performance of crAssphage qPCR Assays 67

4.4.3 Fecal Source Identification and the Human Fecal Viral Metagenome 68

4.4.4 CrAssphage Fecal Source Identification Application 69

5.0 EVALUATION OF CRASSPHAGE-BASED QPCR MARKERS IN AN IMPACTED URBAN WATERSHED 71

5.1 INTRODUCTION 72

5.2 MATERIALS AND METHODS 74

5.2.1 Study Site 74

5.2.2 Sampling and Data Collection 76

5.2.3 Chemical Parameter Characterization 76

5.2.4 Enumeration of Culturable Indicators 77

5.2.5 qPCR Assays 77

5.2.6 Controls 78

5.2.7 Statistical Analysis 79

5.3 RESULTS AND DISCUSSION 79

5.3.1 Site Background 79

5.3.2 Rainfall During the Study Period 80

5.3.3 Chemical Parameters During the Study Period 80

5.3.4 Correlation of Chemical Parameters with Rainfall Events 83

5.3.5 Culturable and qPCR Indicators During Study Period 83

5.3.6 Correlation of Human-associated Markers with Rainfall Events 85

5.3.7 Correlation of chemical and biological water quality parameters 85

5.3.8 Comparison of crAssphage qPCR Assays 87

5.3.9 Site Specific Considerations 88

5.3.10 Environmental Implications 89

6.0 SUMMARY AND CONCLUSIONS 91

6.1 MAJOR FINDINGS 91

6.2 IMPLICATIONS 93

6.3 FUTURE RESEARCH 94

APPENDIX A 95

APPENDIX B 106

APPENDIX C 135

BIBLIOGRAPHY 137

LIST OF TABLES

Table 4.1: CrAssphage qPCR assay oligonucleotides and targeted genomic regions 62

LIST OF FIGURES

Figure 3.1: Percent of sequence reads mapped to the crAssphage genome for metagenomes from different sources……………………………………………………………………….38

Figure 3.2: The percent of sequences mapped for individual samples compared to the fraction of the crAssphage genome covered by the mapped reads.. 39

Figure 3.3: Percent of mapped reads versus position on the crAssphage genome for a) raw sewage from Pittsburgh, PA, USA. b) raw sewage from Shanghai, China c) bat guano from China 40

Figure 4.1: Map representation of the crAssphage genome. 58

Figure 4.2: Abundance of crAssphage and bacterial human-associated qPCR targets in primary influent sewage (Panel A) and environmental water samples (Panel B). 63

Figure 5.1: Depiction of study sampling site within Frick Park. 75

Figure 5.2: Chemical parameters measured along with daily rainfall during the Nine Mile Run sampling period of 30 days. 82

Figure 5.3: Culturable and qPCR indicators along with daily rainfall during the Nine Mile Run sampling period of 30 days. 84

Figure 5.4: Heat map of Spearman's rank correlation coefficients matrix for culturable indicators, qPCR indicators, and chemical parameters.. 86

ACKNOWLEDGMENTS

Many thanks to my family and friends who have supported me throughout my academic career, believing in me when I couldn't believe in myself.

1.0 INTRODUCTION

This dissertation details the development and initial evaluation of novel human-specific, viral-based source tracking technologies. The methods developed in this dissertation research were designed based on a newly discovered bacteriophage, crAssphage, to be used for water management applications.

1.1 MOTIVATION

Evidence exists that enteric viruses are the leading causes of recreational waterborne illnesses, representing a significant portion of the environmental water disease burden [1, 2]. Norovirus is recognized as the single leading cause of outbreaks with total norovirus gastroenteritis incidence amounting to $4.2 billion in direct health systems costs and $60.3 billion in societal costs annually worldwide, with approximately 20 million cases occurring annually in the United States [3-6]. It has been shown that about 16% of all norovirus gastroenteritis illness is caused by environmental sources, such as consuming contaminated water and shellfish [7]. Considering the number of environmentally relevant viruses, total cost and incidence estimates are expected to be much higher; however, the economic and disease burden of enteric viruses is not well characterized. Currently, no environmental water quality regulations exist to measure norovirus or other enteric viruses in environmentally polluted waters [8]. Regulations have been difficult to

implement based on unreliable detection methods for low environmental concentrations of viruses.

Regulations for monitoring environmental waters are currently only available for bacteria. Environmental waters are monitored for fecal pollution by detection of general fecal indicator bacteria (FIB) by culture-based methods. While these methods can give valuable insight into general pollution conditions of bodies of water, many shortcomings remain with these methods. First, these bacteria are general indicators, meaning they are shed by many warm-blooded animals. The animal source of the pollution cannot be ascertained, confounding management efforts. Second, FIB can survive and persist in environmental waters; therefore, their detection is not necessarily indicative of a recent pollution event. Third, detection methods for FIB remain culture-based; however, culture-independent methods are desirable to provide more rapid data acquisition and more immediate regulatory action [9]. Lastly, FIB have been shown to not correlate well with viral health risk. As stated above, viruses represent a significant health burden in environmental waters, with the incidence of viral environmental outbreaks rising and viral outbreaks even occurring when bacterial indicators were undetected or below regulated levels [1, 2]. Current research efforts seek to address these shortcomings using novel technology development.

Microbial source tracking has emerged as a research field in the past ~25 years to determine the sources of fecal pollution to environmental waters. Improving on the lack of specificity of FIB, many assays have been developed to detect fecal material specific to various species of animals including cows, pigs, chickens, dogs, ducks, geese, and gulls [10-14]. In addition, numerous assays have been developed to specifically detect human fecal material [15-17]. Human-specific assays are particularly desirable as human waste has been shown to pose a

greater health risk to humans than waste from other animals [2, 18, 19]. In addition, pollution from human sources is more readily controlled than waste from other animals, such as wildlife.

Several assays for detection of human fecal waste have been developed based on viruses, both pathogenic and non-pathogenic [20-24]. These assays have been devised for use as a direct measure of viral health risk of an impacted water body. Despite their direct correlation with health risk, enteric viruses have proven difficult to track due to lack of culture methods and dilute concentrations in environmental waters [8, 25-28]. Since they are specific to human sources, their detection results in high confidence that human fecal pollution is present. Conversely, as they are not reliably detected at low concentrations, their absence in environmental waters does not instill the same confidence that human fecal pollution is absent. Consequently, there is a critical need for a surrogate indicator of enteric viruses in the environment to accurately assess health risk and guide remediation efforts.

Bacteriophages have long been considered as surrogates for modeling viruses in environmental systems and studies [29, 30]. Bacteriophages are useful due to their similar size and morphology as pathogenic viruses, in addition to their high abundance in sewage [31]. Further, they pose no health risk to human researchers. Recently, a novel bacteriophage, crAssphage, was discovered by metagenomic data mining [32]. The genome sequence for crAssphage was assembled from an individual human gut metagenome and likely represents a crAssphage metapopulation. CrAssphage was found to be more abundant than all other known human gut phages combined and was mostly associated with human gut metagenomes [32]. CrAssphage has the potential to be utilized as a highly sensitive and specific marker of human fecal pollution that could be used as a surrogate of viral-risk in environmental waters. In addition, the initial evidence of high abundance and high human-specificity that crAssphage

exhibits could improve upon shortcomings witnessed with general FIB and other developed MST assays, making crAssphage an ideal target for human-specific marker development.

The research presented in this dissertation investigates the utility of crAssphage as a template for human-specific source tracking technology development. CrAssphage was first evaluated as a candidate for assay development based on bioinformatics data. Next, PCR technologies were developed and evaluated across the crAssphage genome to explore the variable performance across the genome and select the best performing regions. Subsequently, PCR assays were adapted to a qPCR platform and evaluated for performance metrics. Finally, an initial environmental evaluation of the developed crAssphage qPCR assays was performed.

1.2 RESEARCH OBJECTIVES

The research objections of this dissertation have been addressed through the following research projects and represent three manuscripts for publication (represented by chapters 3, 4, and 5). The dissertation is laid out in the following way to present the findings of the research.

Chapter 2: A review of human source tracking technologies: state of the science and research needs

This chapter provides an extensive review of genetic markers that have been proposed and evaluated for detection of human fecal pollution in environmental waters. Discussion is particularly focused on how these technologies have improved on general indicator technologies and shortcomings of the specific assays. This information is synthesized to provide context for how crAssphage markers for source tracking contribute to the tools currently available for managing environmental waters.

Chapter 3: Metagenomic evaluation of the highly abundant human gut bacteriophage crAssphage for source tracking of human fecal pollution

This work has been published as:

Stachler, E. and Bibby, K. Metagenomic evaluation of the highly abundant human gut bacteriophage crAssphage for source tracking of human fecal pollution. *Environ. Sci. Technol. Lett.* 2014, 1 (10), 405-409.

This chapter evaluates the potential of crAssphage for development of human-specific source tracking technologies. An initial metagenomic evaluation is valuable as a first step to determine if more extensive and more expensive lab evaluation is warranted in evaluating potential source tracking microbial targets. In this study, 86 publically available metagenomes were surveyed for crAssphage presence. Metagenomes were taken from various environments, including human sewage, non-human animal fecal matter, and environmental waters. Metagenomic DNA reads were mapped against the crAssphage genome as a proxy for crAssphage abundance in that environment. In addition, metagenome reads from human sewage samples were mapped against other viruses found in the human gut or suggested for source tracking technologies. Results suggested that crAssphage was highly specific to human sewage and more abundant than other viruses typically monitored in sewage, making it an ideal candidate for human source tracking technology development.

Chapter 4: Quantitative crAssphage PCR assays for human fecal pollution measurement

This work has been published as:

Stachler, E.; Kelty, C.; Sivaganesan, M.; Bibby, K.; Shanks, O. Quantitative CrAssphage pcr assays for human fecal pollution measurement. *Environ. Sci. Technol.* 2017, 51(16), 9146-9154.

This work is also the subject of a patent application:

Stachler, E.; Bibby, K.; and Shanks, O. Cross-Assembly Phage DNA Sequences, Primers and Probes for PCR-based Identification of Human Fecal Pollution Sources. Application Number: 62/386,532

This chapter outlines the development of human specific qPCR assays based on crAssphage. These assays were designed and rigorously tested through three main tasks. The first task was to design primer pairs *in silico* across the length of the crAssphage genome. The second task was to evaluate the primer pairs designed in the first task by PCR in the laboratory against fecal reference composite material. The objective of the third task was to adapt the best performing primer pairs based on the results of the second task to a qPCR platform. Two qPCR assays were developed based on crAssphage and were tested head to head with two bacterial qPCR assays for performance criteria including specificity, geographic stability, and an environmental proof-of-concept. This effort resulted in two viral-based qPCR assays that were highly human-specific and abundant in reference materials, having significant potential in water management applications.

Chapter 5: Evaluation of crAssphage-based qPCR markers in an impacted urban watershed

This work is currently in review for publication:

Stachler, E.; Akyon, B.; Aquino, N.; Ference, C.; Bibby, K. Evaluation of crAssphage-based qPCR markers in an impacted urban watershed. In Review.

This chapter provides an initial environmental evaluation of the qPCR assays developed in the research discussed in the above chapter. It is necessary to test the assays in a real-world system to verify assay performance. Stream water was collected daily for 30 days from Nine

Mile Run in Pittsburgh, PA and evaluated for chemical and biological water quality parameters. Biological parameters included culturable bacterial and viral assays as well as bacterial and viral molecular assays. CrAssphage assay occurrence and abundance was correlated with chemical and biological parameters as well as with pollution events. The crAssphage assays were found to correlate with pollution events and other indicators of fecal pollution, both culturable and molecular.

Chapter 6: Summary and conclusions

This chapter summarizes the major findings of the dissertation research, discusses the implications of the research, and suggests areas of future research.

2.0 A REVIEW OF HUMAN SOURCE TRACKING TECHNOLOGIES: STATE OF THE SCIENCE AND RESEARCH NEEDS

2.1 MICROBIAL SOURCE TRACKING INTRODUCTION

2.1.1 Shortcomings of the Current Approach

Environmental waters polluted with fecal matter have been shown to pose a definite health risk to humans [18, 33, 34]. Traditionally, environmental waters have been monitored for fecal pollution with general fecal indicator bacteria (FIB), such as *E. coli* or enterococci. These bacteria were chosen based on their abundance in feces and assumed co-occurrence with pathogens, although these bacteria are not pathogenic themselves. The use of FIB is justified by a correlation with adverse health outcomes at a limited number of epidemiological test sites [35]. Despite this, FIB possess many shortcomings when it comes to environmental water quality monitoring. One limitation of FIB is their potential to persist and grow in environmental reservoirs [36-38]. For instance, FIB have been shown to colonize and grow in river sediments [39]. This suggests their detection may not be indicative of a recent pollution event and inherently limits their utility as an indicator of fecal pollution and subsequent health risk. In addition, FIB are shed by numerous warm-blooded animals, confounding discrimination of the animal source of the pollution [40, 41]. Correct source attribution is crucial to focus monitoring

and remediation efforts where they are most needed. Additionally, although FIB occurrence generally correlates with bacterial pathogen presence in environmental waters, FIB do not always correlate with viral pathogen presence [42-45]. One explanation for this is differential decay rates exhibited by FIB, viruses, and molecular markers in environmental waters [46, 47]. This can lead to FIB not being detected in a system although viruses are still present, giving a false indication of low health risk. Indeed, enteric viral outbreaks in environmental waters have been documented when levels of FIB are below detection limits or regulatory limits [48-51]. Despite obvious shortcomings of FIB for microbial water quality monitoring, this approach persists due to regulatory inertia and lack of viable alternatives, especially for viruses.

2.1.2 Origins of MST

Microbial source tracking (MST) is a recently developed field that seeks to solve one major shortcoming of FIB monitoring by distinguishing the source of fecal pollution in environmental waters. Elucidating the source of fecal pollution is advantageous to inform monitoring and remediation efforts, in addition to inferring health risk. Early MST technology focused on library-dependent methods that compared environmental samples to databases built to characterize the pollution present. Current MST technology uses a library-independent approach that screens for one specific microbe or gene expected to be present in a pollution source. These MST methods operate on the principle that certain strains of microorganisms are host-specific; therefore, their detection informs the animal species contributing to the impairment of the system. Many MST assays have been developed for various animal hosts including wildlife, domestic animals, and farm animals [10-13]. Significant effort has been expended to develop human-specific MST technologies as human waste poses a greater human health risk than waste

from other animals [2, 18, 19, 34, 52]. In addition, human sources of pollution are more readily controlled and managed, such as investing in infrastructure improvements. The field of MST has also focused significant effort on developing culture-independent detection methods, such as qPCR. This allows for technology development that is not limited by culturability of target microorganisms and enables rapid processing of water samples [53]. For the above reasons, this review focuses on molecular MST technologies targeting human fecal pollution.

2.1.3 Terms and definitions

Most technologies discussed in this review are end-point PCR or qPCR assays. End-point PCR is designed to amplify a specific DNA sequence, enabling detection of that sequence in a presence/absence test. Similarly, qPCR is designed to amplify a specific DNA sequence; however, it is capable of quantifying the amount of target DNA present in the sample. A "marker" is defined as the specific DNA sequence that is being detected to indicate the presence of human fecal material. An "assay" is defined as the specific qPCR or end-point PCR experiment that is used to detect the "marker".

MST assays are often evaluated for several criteria to assess the performance of the assays. Most of these criteria are determined by screening assays against target (e.g. sewage) and non-target (e.g. animal feces) DNA. Sensitivity is defined as the percentage of samples expected to be positive (e.g. sewage) that are accurately determined to be positive [54]. Specificity is defined as the percentage of negative samples (samples containing non-target animal fecal DNA) that are accurately determined to be negative [54]. In general, an assay is considered to perform well if both its sensitivity and specificity are 80% or above [55]; however, this limit has been arbitrarily defined and no standard exists for the defined number of samples an assay should be

tested against to determine their sensitivity or specificity. The limit of detection (LOD) describes the smallest amount of marker that may be detected, while the limit of quantification (LOQ) describes the smallest amount of marker that can be accurately quantified [9]. Lastly, markers are often evaluated for geographic stability, in which they are tested against human and non-human fecal DNA distributed across a wide geographic range as certain markers have been shown to vary geographically.

2.2 BACTERIAL MST ASSAYS

The majority of human MST assays have been developed on bacterial targets, specifically the 16S rRNA gene of *Bacteroides* species. *Bacteroides* is one of the predominant genera in the human gut microflora, making it a potential target for development of human-specific markers [56, 57].

2.2.1 *Bacteroides* 16S rRNA Assays

The most common target for human MST assay development to date has been the 16S rRNA gene of *Bacteroides* species. Many of the DNA sequences targeted by the assays discussed in this section overlap in the same genomic region of the 16S rRNA gene of *B. dorei* [9]. Designed assays exhibit differences in performance, which may reflect varied optimization of PCR reaction conditions, specific location of primers, or fecal libraries the assays have been screened against.

2.2.1.1 HF183

HF183, one of the first human source tracking PCR assays designed, has consistently been shown to be among the top-performing MST technologies available. The HF183 assay was first designed as an end-point PCR assay and demonstrated 84% sensitivity when tested against human fecal samples (n=13) but 100% sensitivity when tested against sewage (n=3) [58]. The assay exhibited 100% specificity against an animal fecal library (n=46) and an LOD of 0.01 ng DNA which corresponded to a high 10^5 gene copies [58]. The assay was later adapted to a qPCR assay utilizing SYBR Green chemistry which exhibited 86% sensitivity against human fecal samples (n=7), 100% sensitivity against sewage samples (n=4), and 95% specificity (n=19; cross-reacting with a single chicken sample) [59]. The LOQ of the assay was 5.67 $\log_{10}$ copies/L in freshwater [59]. The HF183 assay was further adapted to TaqMan qPCR chemistry which exhibited 100% sensitivity based on a human fecal composite (n=1) and sewage samples (n=14); however, the assay exhibited 60% specificity, cross-reacting at lower concentrations with chicken and dog composite samples [60]. The HF183 TaqMan assay was shown to form nonspecific amplification products and was therefore modified by replacing the reverse primer with a more specific reverse primer (BacR287) [17]. The two assays were then tested head-to-head and exhibited the same sensitivity and specificity and similar concentrations in all reference samples; however, the HF183/BacR287 assay eliminated the nonspecific amplification products experienced with the former assay which improved the limit of detection of the assay slightly [17]. Both the original TaqMan assay and the new HF183/BacR287 assay demonstrated 67% sensitivity with human fecal samples (n=6), 100% sensitivity in sewage (n=58), and 93% specificity (n=123, cross-reacting with chicken and turkey samples) [17]. Again, both assays exhibited an LOD of 1 pg DNA per reaction when human fecal DNA was used, but the

HF183/BacR287 assay showed a slight improvement with an LOD of 5 pg DNA per reaction versus 50 pg DNA per reaction for the former assay when sewage DNA was used for input [17].

The HF183 assays have been included in several comparison studies between other developed assays and have consistently demonstrated the best performance or among the best performances [60-64]. A multi-lab study that investigated 23 human-specific assays found that the only two assays to exhibit both a sensitivity and specificity of 80% or above were the HF183 endpoint and HF183 SYBR assays [55]. Due to their consistently high performance, the HF183 assays are widely considered to be top performing technology in human MST methods. However, the assays have also shown cross-reactivity with non-target animal fecal DNA, namely from cat, dog, and chicken samples [17, 59, 61, 65].

2.2.1.2 BacH

In addition to HF183, multiple other MST assays have been designed from the 16S rRNA gene of *Bacteroides* species. BacH is a TaqMan qPCR assay that upon initial development exhibited a 95% sensitivity in human fecal samples (n=21) and 100% sensitivity in sewage (n=20) [66]. The assay exhibited 99% specificity (n=302), cross-reacting with a single cat fecal sample, and was detected down to 6 copies per reaction [66]. The assay performed well when compared head-to-head with other MST assays in a single lab study, displaying 100% sensitivity and 94% specificity, cross-reacting with sheep, goat, and dog samples [63]. A multi-lab comparison study found the assay to exhibit a lower specificity at ~75% [55]. The assay was also evaluated for international stability against samples from 16 countries across 6 continents. This study demonstrated the importance of local validation of assays, as the global sensitivity of BacH was 77% based on human fecal samples (n=61) and specificity was low at 53% (n=219) [67].

2.2.1.3 HuBac

Another qPCR TaqMan assay developed on the 16S rRNA *Bacteroides* gene is HuBac. This assay was found to have 100% sensitivity based on human fecal samples (n=3); however, it had a specificity of 68% (n=18), cross-reacting with pig samples [68]. The low specificity of this assay has been consistently verified since its initial development, exhibiting specificities from 22.7-63% in studies that evaluated HuBac head-to-head with other assays, despite 100% sensitivities in all studies [61, 63, 65].

2.2.1.4 BacHum

BacHum, another TaqMan qPCR assay based on the 16S rRNA gene of *Bacteroides*, demonstrated high initial performance, exhibiting a 67% sensitivity in human fecal samples (n=18), a 100% sensitivity in sewage (n=14), and a 98% specificity (n=41, cross-reacting with one dog sample) [65]. It also exhibited a low LOD of 3.5 gene copies per reaction and an LOQ of 30 gene copies per reaction [65]. The high performance was verified in a follow-up study, displaying 100% sensitivity (n=50); however, the assay did cross-react with pig, sheep, horse, and dog samples, although the overall specificity was 96% [63]. As is the case with many MST markers, initial evaluations are favorable but performance decreases as additional studies are conducted. Another study confirmed the assay's high sensitivity at 100% in sewage samples (n=12) but found a much lower specificity of 71%, cross-reacting with cow, horse, and dog samples [69]. A multi-lab comparison study demonstrated very high sensitivities (>90%) but showed the specificity to vary widely between labs (0-60%) [55]. Similarly, in a global study, the assay exhibited 87% sensitivity in human fecal samples (n=61) and 68% specificity (n=219) [67]. In addition, studies can have seemingly contradictory results based on the fecal libraries screened. One study found the BacHum assay to be 100% specific (n=25); however, the assay

was detected in 25% of human fecal samples (n=12) and 0% of sewage samples (n=5) [62]. The assay was only detected in human samples, which led to the high reported specificity but was not found in all or even most of the human-derived samples. The lack of detection in human samples likely contributed to the reported high specificity value, as human-specific assays are typically found at concentrations orders of magnitude lower in non-target animal samples and therefore were most likely below the limit of detection in the animal samples in this study.

2.2.1.5 HumanBac1

Another 16S rRNA *Bacteroides* marker, HumanBac1, was found to exhibit 100% sensitivity (n_{fecal}=4) but low specificity at 10% (n=10), cross-reacting with cow and pig samples at similar concentrations to human samples [70]. The assay did exhibit a low LOD and LOQ, both at 4.3 copies per reaction [70]. Further evaluation of the assay demonstrated continued problems with low specificity. Comparison studies including the assay demonstrated specificities of 57% (n=30) and 79% (n=136), cross-reacting with cow, sheep, horse, dog, and kangaroo samples [63, 71].

2.2.1.6 BacHuman

The BacHuman qPCR assay again targets the 16S rRNA gene of *Bacteroides*. Initial evaluation found the assay to be 100% sensitive in sewage samples (n=16) and 81% specific (n=54), cross-reacting with pig, dog, and cat samples [72]. The assay exhibited a low LOD at 6.5 gene copies per reaction; however, demonstrated poor correlation with FIB in river water samples [72].

2.2.2 Other *Bacteroides* Assays

Bacteroides 16S rRNA genes have been targeted for assay development due to their high abundance, which facilitates detection. However, since the gene is highly conserved it often cross-reacts with non-target samples. This has led to efforts to target other genes in *Bacteroides* species that may be more specific to humans. Potential human-specific sequences were found using genome fragment enrichment and used to design PCR assays based on a region of a hypothetical protein and on a putative RNA polymerase sigma factor [15]. The assays were found to be 100% sensitive based on sewage samples (n=16) and 100% specific (n=90) [15]. Based on these results, the assays were adapted to qPCR assays as HumM2 and HumM3. Both assays exhibited 100% sensitivity (n_{fecal}=16; n_{sewage}=20) while HumM2 exhibited a slightly higher specificity (n=265) at 99.2% compared to 97.2% for HumM3 [16]. A study comparing multiple MST assays found the assays to have some of the highest specificities at 90.9% and 95.5% (n=158), cross-reacting with sheep, elk, and chicken samples [61]. A multi-lab study found that HumM2 was >80% sensitive but was <80% specific [55]. In addition, the HumM2 assay was found to be significantly less reproducible than other qPCR assays such as HF183 and BacHum, due to lower concentrations in test samples [64].

Another assay targeted the gyrB gene of *B. fragiles*, a gene involved in super-coiling of DNA. The assay was found to be 100% sensitive based on human fecal samples (n=10) and 97% specific (n=30) [71]. Despite the high performance demonstrated in the initial study, an additional comparison study showed the assay to have exhibit lower performance at ~70% sensitivity and ~90% specificity [55].

Another qPCR assay was designed on the α-1-6, mannanase of *B. thetaiotomicron*, a dominant species in the human gut microbiome [73]. The assay exhibited high initial

performance criteria with 100% sensitivity (n_{fecal}=10; n_{sewage}=20) and 100% specificity (n=160) [74]. The assay exhibited 100% sensitivity in another comparison study, but only ~50% specificity [55].

2.2.3 Other Bacterial Targets

Other bacterial targets for MST technology development have been investigated. The 16S rRNA gene of *Bifidobacterium adolescentis* has been targeted as a marker for MST method development. The sensitivities of the assays ranged from 90-100% while the specificities ranged from 89-100% [75-77]. However, when tested head-to-head the HF183 assay was shown to perform better than the *Bifidobacterium* assay [75]. Additionally, the enterococcal surface protein (esp) of *E. faecium* has been targeted for MST method development. The sensitivities of the developed assays have been reported from 67-100%, while the specificities have been reported from 95-100% [78-80]. However, these uncommon methods are often left out of comparison studies and must be further verified to accurately assess performance metrics.

2.2.4 Challenges and Limitations

Many human MST assays have been developed based on bacterial targets, specifically the 16S rRNA gene of *Bacteroides* species. Performance varies widely between the assays discussed, and assay performance often varies widely between studies and fecal libraries. However, when looked at as a whole several trends appear. Overall, bacterial MST assays exhibit generally high sensitivities, indicating their high concentrations in human fecal material and sewage and facilitating their detection in diluted environmental waters. While the assays often exhibit high

specificities, no assay has yet been found to be 100% specific towards humans in all fecal libraries. The assays often cross-react with a wide range of non-target animal sources, and therefore may be less useful in areas where these animal species are frequently found. Between several comparison studies, the marker that consistently performs the best is HF183, and therefore this assay has been developed the furthest. Some studies have even been conducted to relate HF183 marker presence to predicted health risk of impaired waters [34, 81]. One study found that a gastrointestinal illness rate of 30 per 1000 swimmers occurred at HF183 concentrations of 4200 copies/100 mL of recreational water [81]. This number changes, however, if other sources of pollution are present. It was estimated that for risk attributed to gull sources at California beaches, HF183 would have to be present at 12-630 copies per 100 mL to not exceed the same level of risk [34]. However, these studies were based on models and not epidemiological data. In addition, research has been conducted to correlate bacterial MST markers with general indicator and viral pathogen presence and decay in environmental waters; however, the results vary widely, finding no correlation with pathogen presence in some studies while finding positive correlation with pathogen presence in other studies [44, 47, 75, 82-85]. Additional studies are necessary to elucidate true trends between bacterial MST markers and health risk, especially health risk from viral pathogens.

2.3 VIRAL MST ASSAYS

Enteric viruses are a significant source of health risk from environmental waters polluted by human fecal waste [1, 2]. They can be transmitted directly to the environment through combined sewer overflow events, leaky septic systems, or illegally connected downspouts. In addition,

viruses can be released from wastewater treatment plant effluent, where viruses pass through at higher rates than FIB [86, 87]. As discussed above, current methods of monitoring using FIB or bacterial qPCR assays do not necessarily correlate with viral pathogen presence in environmental waters. It is estimated that exposure to coastal waters results in millions of excess GI illnesses in the U.S. and hundreds of millions of excess GI and respiratory illnesses globally every year [88, 89]. However, the actual human health risk due to virus exposure in environmental waters and associated costs have not been characterized even though the percentage of outbreaks attributed to viruses have been rising in recent years [1]. Due to this, MST technologies have been developed based on both pathogenic and non-pathogenic viruses in order to better capture viral risk in environmental waters.

2.3.1 Enteric Pathogen Viral Assays

Many qPCR assays have been developed based on pathogenic enteric viruses. These assays have typically been developed as diagnostic tools and not necessarily for source tracking applications. These viruses have also been explored in a source tracking context as improved human-specific markers and to directly measure the health risk of exposure to an impacted water body. Advances in metagenomics may make it possible to screen for all pathogens of concern in the future [90]; however, screening for every virus is a currently laborious practice and therefore the most relevant viruses are often evaluated.

2.3.1.1 Norovirus

Norovirus is recognized as the leading cause of viral health risk in environmental waters. Norovirus is the cause of more than 50% of all outbreak-associated gastroenteritis and more than

90% of non-bacterial outbreaks [5]. In addition, norovirus has been shown to be the cause of 45% of all viral caused outbreaks in recreational waters [1]. Due to the significance of norovirus, PCR assays have been developed for its detection [22, 91-93]. PCR assays have successfully detected norovirus in 99% of human fecal samples that were shown to be positive for norovirus by microscopy [93]. One study developed norovirus assays for specific use in source tracking applications. The assays for norovirus GI and GII exhibited sensitivities of 40 and 67% in human fecal samples (n=15), respectively, and each exhibited 82% sensitivity in sewage (n=11) [91]. The assays also demonstrated 100% specificity (n=56) [91]. A multi-lab study comparing different viral assays demonstrated that norovirus-based assays exhibited 0% sensitivity but 100% specificity [94]. However, it is generally accepted that norovirus assays are themselves not ideal MST technologies due to the well-documented seasonal variation in norovirus abundance throughout the year based on active human infection rates, with peak concentrations occurring in winter months [92, 95, 96]. This contributes to the lack of data on norovirus for source tracking applications.

2.3.1.2 Adenovirus

Adenoviruses have been shown to be the second leading cause of viral outbreaks from exposure to contaminated environmental waters [1]. Several end-point PCR and qPCR assays for the detection of human adenovirus have been developed; however, these assays were not specifically developed for source tracking technologies and therefore do not report sensitivity or specificity estimates [20, 21, 97, 98]. One study developed adenovirus assays as part of a toolbox approach for source tracking applications. This study found the assays to exhibit 0-7% sensitivity in human fecal samples (n=15), but 36-100% sensitivity in sewage samples (n=11) [91]. The assays were also 100% specific (n=56) [91]. A multi-lab study evaluated adenovirus assays next to

other MST technologies and demonstrated assay sensitivities from 0-16.7%, while specificity was 100% [94]. In addition, adenovirus assays, along with human polyomavirus assays, have been demonstrated to be suitable indicators of norovirus presence in polluted waters while being present at higher concentrations and not exhibiting seasonal variability [95].

2.3.1.3 Enterovirus

Enteroviruses, including echovirus and poliovirus, have also been shown to cause outbreaks in environmental waters [1]. Due to their health relevance and presence in environmental samples, several PCR assays for enterovirus have been developed [99-102]. Again, these assays were developed as diagnostic tools and not for specific source tracking applications, so knowledge of their occurrence and abundance in a range of human and animal fecal samples is limited. One multi-lab study included enterovirus assays in their comparison of MST technologies, finding the sensitivities of the assays ranged from 0-13.2% while the specificities ranged from 84.6-100% [94].

2.3.2 Non-Pathogenic Viral Targets

Pathogenic viruses are often not present at high enough concentrations to be accurately detected as seen in the discussion of the above assays. In addition, it is not economical and feasible to screen for all pathogenic viruses that could potentially impact a body of water, due to the number of unique viruses of concern and the portion of the viral human gut microbiome that remains uncharacterized due to lack of similarity to currently described viruses [103]. In addition, the detection of these viruses can vary throughout the year based on the active human infection rate in a given population. Therefore, non-pathogenic viruses have been proposed as indicators for

viral pathogen presence. An appropriate viral indicator would co-occur with viral pathogens, be present at higher concentrations than viral pathogens for ease of detection, be consistently shed by human populations, and decay at a similar or slower rate to viral pathogens to serve as a conservative indicator.

2.3.2.1 Human Polyomavirus

Human Polyomavirus (HPyV) has been shown to cause lifelong asymptomatic infections in much of the human population, and is primarily shed in urine [104]. Although it is typically considered nonpathogenic, it is known to cause kidney nephritis in the immunocompromised [105]. Due to its wide presence in the human population, HPyV has been developed as a target for human source tracking applications. An end-point PCR assay was designed based on the T antigen gene and found to be 100% sensitive in human sewage (n=36) and 100% specific (n=25) [106]. The assay was adapted to a qPCR chemistry and found to be 100% sensitive towards sewage samples (n=39) and 23% sensitive towards individual urine samples (n=26) [23]. The assay did not cross-react with any non-target animal fecal (n=117) or urine (n=10) samples and was detected down to 10 copies per reaction [23]. A multi-lab study again found the assay to be 100% specific, but sensitivities were much lower at 7.9-10.5% [94]. HPyV has been tested head-to-head with the bacterial source tracking marker HF183 (detailed in section 2.2.1.1) and consistently exhibited a higher specificity at 100% compared to 81-96% for HF183 [23, 107-109]. However, HPyV is less abundant in reference materials than HF183, showing concentrations 2 orders of magnitude lower [107, 108]. HPyV has been shown to not correlate with indicator bacteria [106], but has been shown to correlate with adenovirus in presence and decay studies [23, 110].

2.3.2.2 *Pepper Mild Mottle Virus*

Pepper mild mottle virus (PMMoV) is a plant pathogen that is widespread and abundant in human fecal samples [111]. PMMoV is a member of the genus tobamovirus, capable of infecting a wide variety of pepper plants [112]. PMMoV has a single stranded RNA genome that was first sequenced in 1991 [113]. The virus is produced in copious amounts in the plants it affects and is assumed to be present in human sewage due to human consumption of pepper products such as hot sauce [112]. Since this marker is dietary in origin, it does not depend on active human infection rates. A qPCR assay was designed for its detection and demonstrated 67% sensitivity in human fecal samples (n=18), 100% sensitivity in sewage (n=12), and an LOD of 100 copies per reaction, but showed imperfect specificity cross-reacting with chicken and gull samples [24, 111]. A recent study tested PMMoV head-to-head with HF183 in samples from Costa Rica and found the PMMoV assay to perform better, with 100% specificity compared to 94% for HF183 [109]. Another qPCR assay designed on PMMoV was found to be 95% sensitive in human fecal samples (n=20) and 83% specific (n=53; cross-reacting with chicken, cow, and geese samples) [114]. This same study looked at survival of PMMoV, HPyV, and adenovirus spiked into river water. HPyV decayed the fastest, while PMMoV exhibited the highest stability [114]. In seawater exposed to wastewater treatment effluent, PMMoV co-occurred with several viral pathogens, including norovirus [24]. PMMoV has also been shown to exhibit increased persistence through wastewater treatment than pathogenic viruses, suggesting it could be a conservative indicator of enteric viruses [86].

2.3.3 Challenges and Limitations

Pathogenic human viruses are very specific towards human sources; however, they often are not present at high enough concentrations to be accurately detected. Because of this, their detection in impacted waters results in high confidence that human fecal pollution is present. Alternatively, their absence does not instill high confidence that human fecal pollution is absent. A multi-lab study found that enterovirus and adenovirus assays exhibited high false negative rates in test samples spiked with human fecal or sewage material (i.e. the assays were not detected although human fecal contamination was present). The false negative rates for human fecal samples were 100% and 90% for enterovirus and adenovirus, respectively, and 60% and 40% in samples spiked with sewage [115]. Other viral targets have been suggested to act as indicators for viral pathogens. PMMoV is highly abundant and therefore easy to detect but shows some cross-reaction with animals that may obtain the virus in their gut via their diet. Human Polyomavirus has yet to cross-react with any non-target animal sources and is detectable in environmental waters. Despite this, viral markers are often detected at concentrations 2-4 orders of magnitude lower than bacterial indicators [107, 108, 116, 117]. This could result in false negative results when monitoring sites that are subject to low levels of fecal contamination.

Due to their lower abundance, water samples are often concentrated before screening with viral-based source tracking assays. However, concentration methods vary widely in their efficiencies and recovery rates. For example, enteroviruses were recoverable at 51% in freshwater and 23% in seawater using negatively charged cellulose acetate/nitrate filters [100]. Another study found 30-31% recovery of adenoviruses and 40-78% recovery of polyomaviruses when directly extracting from a negatively-charged filter after acidification [28]. The same study found only 2.4-5.3% recovery for adenoviruses and 5.9-12% recovery for polyomaviruses using

an absorption/elution procedure [28]. In addition, PCR inhibitors can be co-concentrated along with viruses, making it difficult to retain accurate results [28]. Many of the technologies available are also expensive, making them impractical for wide scale sampling of environmental waters [27]. To improve confidence in viral-based assays, more reliable, efficient, and cost-effective concentration methods are needed.

2.4 BACTERIOPHAGE MST ASSAYS

Bacteriophages, viruses that infect bacteria, have long been investigated as surrogates for modeling pathogenic virus' behavior in environmental fate, decay, and disinfection studies. Bacteriophages, or "phages", have similar size and morphology to viruses and are suggested as surrogates due to ease of detection, cost, abundance in environmental waters, and safety compared with viral pathogens [30]. Bacteriophages have been shown to be more abundant than bacteria in most environments, which may ease dilution problems experienced with monitoring viruses [31]. Despite being more abundant than bacteria, they generally remain more difficult to detect as bacteria are easier to concentrate based on size exclusion filtering. Bacteriophages are safer to study than pathogenic human viruses, as phages do not infect humans. Likewise, facilities that study pathogenic viruses are costlier to set up and maintain compared to labs that research phages that can be studied in biosafety level 1 (BSL1) facilities. All of these reasons, along with their similarity to viruses in size and shape, make phages potentially ideal targets for human MST technology development. While this review focuses on molecular techniques of detection, many of the assays for detection of bacteriophages for source tracking remain culture based and will be discussed below.

2.4.1 F^+ RNA Coliphage

Much of the bacteriophage MST technology development has targeted coliphages. Coliphages are bacteriophages that specifically infect *E. coli*. There are distinct types of coliphage characterized by the means in which they infect *E. coli*. Somatic coliphages infect *E. coli* via their outer membrane and F^+ coliphages infect *E. coli* via the pilus appendage found on the surface of *E. coli* cells [118]. Somatic coliphages are considered general indicators of fecal pollution since they are not specific to human fecal sources and are thus not discussed in this review. However, one method for source tracking of human pollution detects the frequencies of the four F-specific (F^+) RNA coliphage genotypes to determine if the contamination has a human source. This method works on the basis that groups II and III are predominant in human sources and groups I and IV are predominant in animal sources; therefore, genotyping the phages allows the source to be discerned [119]. Oligonucleotide probes were first used to genotype the F^+ RNA phages, and correctly identified the source for 100% of samples (derived from surface waters, sewage effluent, and animal fecal samples) [119]. Two different groups developed qPCR assays to simultaneously detect all four subgroups with an LOD of 10 copies per reaction, although sensitivity and specificity were not measured [120, 121]. A subsequent study that used F^+ RNA phage genotyping as part of a toolbox approach showed that the assays were not specific to the expected source. Group II and III phages (expected to be of human origin) were not detected in human fecal samples (n=15) [91]. The only group detected in human fecal samples was group I, detected in 7% of the samples [91]. In addition, group II phages were found in 80% of pig samples, 100% of deer samples, 50% of cow samples, and 6% of duck samples [91]. However, groups II and III were detected more frequently in human sewage samples (n=11) at 91% and 100%, respectively, compared to 55% for group I and 18% for group IV [91]. Another set of

qPCR assays was developed and were not tested for specificity; however, in human sewage (n=7) groups I and II were detected in 100% of samples, group III in 86%, and group IV in 0% [122]. A multi-lab comparison study included the assay for group II F^+ RNA phages and found the assay was 18.4% sensitive and 84.6% specific, cross-reacting with gull and pigeon samples [94].

2.4.2 Phages of *Bacteroides*

Many assays for bacteriophage detection for source tracking remain culture based. A large portion of these methods focus on culturing phages that infect *Bacteroides* hosts. Just as *Bacteroides* species have long been pursued for human source tracking due to their abundance in the human gut, *Bacteroides* bacteriophages have also long been suggested [29, 123, 124]. *Bacteroides* phages generally have a narrow host range, do not replicate in the environment since their hosts are anaerobic, and do not exhibit seasonal variability [124]. These phages are also more persistent in the environment and more resistant to wastewater treatment than FIB, similar to pathogenic viruses and other bacteriophages [124]. A method for detecting phages of *Bacteroides* GB-124 was 100% sensitive in sewage samples (n=110) and 100% specific (n=30) [125]. A follow-up study confirmed the assay's 100% sensitivity (n_{sewage}=12) and 100% specificity (n=19) and showed the phages to co-occur with norovirus and adenovirus in sewage samples [126]. In contrast, another study found GB-124 bacteriophages had the highest sensitivity of four human-specific bacteriophage methods tested at 66.7%, but had the lowest specificity at 57.7%, cross-reacting with all animal samples except for deer and goose [94]. In addition, it may be necessary to isolate a new host for various geographic regions, and these bacteriophage methods rely on plaque assays cultured in anaerobic conditions, so results cannot

be obtained same-day [127]. Despite this, it has been speculated that sequence based technologies may identify more abundant phages of the non-culturable *Bacteroidetes* which appear to outnumber the culturable *Bacteroidetes* [124].

2.4.3 Challenges and Limitations

Technologies targeting bacteriophages for source tracking of human fecal pollution have faced many challenges that have limited their application in water management. A major problem with existing bacteriophage assays is their lack of human-specificity. Since phages occupy the same habitats of their bacterial hosts, they will cross-react with other animal sources if their bacterial hosts are not specific towards humans. As was discussed in the above bacterial section, many of these assays have also demonstrated cross-reactivity with animal sources; therefore, it should not be surprising that phage assays cross-react as well. While phages can be more abundant than pathogenic viruses in environmental samples, their lack of specificity remains a significant hurdle. In addition, many of the methods developed still rely on culturing phages for enumeration, resulting in longer delays in data acquisition. It could be advantageous to further investigate the development of qPCR assays based on bacteriophages for source tracking applications.

2.5 RESEARCH NEEDS

The review of MST technologies above demonstrates the number of assays available and the varied performance exhibited by the assays. However, more research is necessary before these

technologies can replace FIB for regulation of environmental water quality management. The assays discussed above often exhibited a wide range of performance across studies that used different fecal libraries. Standards for evaluation of new assays are necessary to directly compare MST technologies to each other. Additional multi-lab studies should also be conducted evaluating multiple MST technologies to eliminate lab-to-lab variation. In addition, many studies have evaluated sensitivity and specificity for MST markers; however, there is a dearth of studies correlating markers to pathogens and health risk. In order to understand if we are properly capturing viral health risk with these technologies, additional studies are necessary. Assays that correlate with pathogen presence should next be evaluated in epidemiology studies to investigate if these assays can likewise be correlated with realized health risk. Ultimately, a tool-box approach may be necessary in which environmental waters are monitored with several markers (a combination of bacterial and viral) in order to capture a complete picture of total health risk. In addition, novel technologies should continue to be pursued as more of the human microbiome is explored through bioinformatics approaches.

2.5.1 Ideal Marker Characteristics

The above discussion shows the vast range of methods for testing and verifying MST assays. Assays based on bacteria, viruses, and bacteriophages all have advantages over the other types of assays. Likewise, all existing assays exhibit at least one shortcoming. Bacteria-based assays are abundant in reference materials and are easy to detect; however, these assays commonly cross-react with fecal DNA from non-human animal sources and are not representative of viral pathogen health risk. Viral-based assays are highly specific towards humans and are a more direct measure of health risk; however, they are present at lower concentrations in reference

samples, lowering confidence in these tests when they fail to be detected. Bacteriophage-based assays are more representative of viruses in environmental systems; however, they also commonly cross-react with non-human animal samples and detection methods remain mostly culture based.

Novel MST methods continue to be developed in the hopes of designing a marker that can eliminate the obstacles experienced with the above assays. Examining existing technologies highlights several characteristics that an ideal marker would exhibit. First, an ideal marker of human fecal pollution would be present in all target reference materials, such as sewage (100% sensitive). Not only would an ideal marker be present, but it should be abundant in human reference materials to ease its detection. The marker would likewise be absent from all non-target animal feces (100% specific). The assay should also have a known geographic characterization, as assays have been shown to perform differently on a regional or global scale. The assay should also correlate with pathogen presence and survival in the environment. Finally, the ideal marker detection method should be rapid (likely based on molecular methods) to facilitate water quality management decisions. A marker that exhibits all of these characteristics would greatly improve the way environmental waters are monitored and managed to protect public health.

3.0 METAGENOMIC EVALUATION OF THE HIGHLY ABUNDANT HUMAN GUT BACTERIOPHAGE CRASSPHAGE FOR SOURCE TRACKING OF HUMAN FECAL POLLUTION

This work has been published as:

Stachler, E. and Bibby, K. Metagenomic evaluation of the highly abundant human gut bacteriophage crAssphage for source tracking of human fecal pollution. *Environ. Sci. Technol. Lett.* 2014, 1 (10), 405-409.

Recently, a highly abundant and widespread bacteriophage, named crAssphage, was identified in the human gut. Here, 86 publically available metagenomes were surveyed to determine the presence and abundance of crAssphage in various environments, and to identify its utility for source tracking of human fecal pollution. CrAssphage was found to be highly abundant in sewage and biosolids from the United States and Europe, and less abundant in sewage from Asia and Africa. CrAssphage was not definitively identified in other samples, including animal fecal material, with the exception of bat guano. Approximately half of mapped reads in the bat guano metagenome clustered to orf00045 in the phage genome, suggesting homology to a closely related phage. These results indicate the potential utility of a crAssphage based marker for source tracking of human fecal waste and highlight the utility of metagenomic approaches for initial identification and verification of microbial source tracking markers.

3.1 INTRODUCTION

Markers of fecal pollution in the environment are used to protect public health and regulate recreational waters, drinking water resources, and food production [128]. Fecal indicator bacteria (FIB) are generally used to regulate and test for sewage contamination in natural waters, despite problems affecting their usefulness as indicators of fecal pollution, including challenges with indicating viral pathogen presence and source determination [9]. In response to the many problems surrounding FIB, significant effort has been expended to identify other potential markers for microbial source tracking (MST). A focus has been on culture free methods such as quantitative PCR (qPCR) to allow reliable same day contamination information [9, 64]. There is especially high value in identifying human specific markers since human waste poses a greater risk to human health than most other animal sources [18, 128]. Several markers have been developed and tested for various animals, including humans, enabling tracking of fecal contamination back to its source [10, 14]. Recently, considerable effort has been expended to identify the best-performing MST markers and methods [9, 55], including evaluation of viral and bacteriophage markers [94]. The ideal human waste marker would be both highly specific and sensitive, and have well-characterized geographic variability [9].

Viruses, including bacteriophages, have been proposed to be a better indicator of fecal pollution than FIB [24]. Bacteriophages of the bacterial genus *Bacteroides* have long been suggested for source tracking [29]. In contrast to other bacteriophage groups proposed for source tracking (e.g. coliphages), *Bacteroides* phages are not found in pristine waters and cannot replicate in the environment [29]. *Bacteroides* phage GB-124 was tested as a fecal indicator and found to be human specific when tested against many animal samples; however, it was found in lower concentrations than FIB, coliphages, and adenovirus, complicating detection when diluted

in natural waters [129]. Thus, alternative viral markers are an active topic of investigation. For example, pepper mild mottle virus (PMMoV), a plant pathogen found in food and in high numbers in sewage, has been suggested as an indicator since it is independent of active human infection rates [24]. PMMoV has also been found to be more abundant than other viruses in raw sewage, including adenovirus, Torque teno virus, norovirus, and polyomavirus [24]. In addition to the use of viruses or other pollution markers, the culture-independent sequencing of environmental samples, or 'metagenomics', has been proposed for the identification of both pathogens and indicators [90, 130, 131].

Recently, a novel bacteriophage was identified in the human gut [32]. The genome of the previously unrecognized "crAssphage" was produced through metagenomic data mining and verified by Sanger sequencing. CrAssphage was found to be more abundant than all other previously identified phages in the human gut combined, and to be conserved across individuals [32]. CrAssphage was suggested to be a *Bacteroides* phage using co-occurrence profiling [32]. Initial evaluation indicated that crAssphage was most common in fecal metagenome samples [32]. Based on the great abundance of crAssphage in the human gut, compounded with *Bacteroides* phages previously being suggested for source tracking [29], we hypothesize that crAssphage has significant potential as a highly sensitive and specific MST marker for human fecal contamination.

In the current study, we seek to evaluate the utility of crAssphage as an MST marker by searching existing viral and microbial metagenomes from a wide variety of environments (sewage, biosolids, terrestrial animals, fish, reclaimed water, freshwater, hypersaline water, marine water, mosquito, coral, microbialites) to identify the presence and abundance of crAssphage in those environments. In addition to exploring the utility of crAssphage as an MST

marker, this study serves to provide an example of how existing metagenomic data can be utilized for initial identification and evaluation of MST markers.

3.2 MATERIALS AND METHODS

3.2.1 Genome downloads

The crAssphage genome was downloaded from Genbank (accession JQ995537) [32]. Other viral genomes downloaded from the NCBI nucleotide database include pepper mild mottle virus (accession NC_003630.1) [113], human adenovirus F (accession NC_001454) [132], human polyomavirus BK (accession NC_001538) [133], Torque teno virus (accession NC_002076) [134], and norovirus GII (accession HQ449728) [135].

3.2.2 Identification of metagenomes

Google Scholar and NCBI Pubmed were used to search for existing viral and microbial metagenomes of various environments. Both viral metagenomes and whole microbial metagenomes were utilized. For the purposes of this study, viral metagenomes were considered to be metagenomes where virus-like particles were concentrated prior to sequencing, and all other metagenomes were considered to be whole microbial metagenomes. Targeted sequencing approaches were not considered. The metagenomes found in the literature search contained raw sewage and anaerobic digester samples from the United States (n=12) [103, 136, 137], a raw sewage sample from Europe (n=1) [136], raw sewage samples from Africa (n=2) [103, 136], raw

sewage samples from Asia (n=5) [103, 138], reclaimed water (n=6) [138, 139], viral (n=5) and whole microbial (n=8) metagenomes from various terrestrial animals (cow=4 [140], chicken=2 [140], pig=1 [141], bat=2 [142, 143], sea lion=1 [144], rodent=3 [140, 145, 146]), freshwater (n=6) [140, 147], fish (n=2) [140], mosquito (n=3) [140], hypersaline water (n=12) [140], marine water (n=9) [140, 148, 149], coral (n=6) [140], and microbialites (n=3) [140, 150]. The accession numbers from these projects were used to download the metagenomic datasets from either the NCBI Sequence Read Archive (SRA) or MG-RAST [151], or obtained directly from the author. Since there was an extremely large number of reads per sample from the sewage sludge viral metagenomes [137], a random subset of reads was compared to the crAssphage genome. In addition, the NCBI SRA and MG-RAST were searched for additional relevant metagenomes not found through the literature search. Metagenomes found this way include raw sewage from Shanghai, China (n=1), bovine rumen (n=1), chicken cecum (n=2), and canine feces (n=2). Accession numbers for all metagenomes used in this study, along with additional sample details and alignment results, can be found in Table A1.

3.2.3 Alignment to the reference genomes

Metagenomic sequences were mapped onto viral genomes in the CLC Genomics Workbench 7.0.3 (CLC Bio, Aarhus, Denmark) using default parameters (no masking, mismatch cost – 2, insertion cost – 3, deletion cost – 3, length fraction – 0.5, similarity fraction – 0.8, non-specific match handling – map randomly). Coverage data was computed for samples with at least 0.01% of reads mapping to the crAssphage genome.

3.3 RESULTS AND DISCUSSION

Metagenomic sequences, termed 'reads', from various environments were mapped against the crAssphage genome to survey these environments for the presence and abundance of crAssphage. Figure 3.1 shows the percentage of reads mapped to the crAssphage genome for each environment, averaged over the total number of reads from each environment. Of the 86 metagenomes surveyed, 51 metagenomes had at least one read that mapped to the crAssphage genome. However, only 17 metagenomes had at least 0.01% of the total reads map to the crAssphage genome. All metagenomes with less than 0.01% of total reads mapping had only 1-16 reads map (0.00017-0.0048% of total reads), and were excluded from further analysis due to lack of genome coverage. The mosquito, coral, and microbialites environments (Table A1) did not include any samples with at least 0.01% of reads mapping and are not included in Figure 3.1. Samples that had at least 0.01% of reads mapped to the crAssphage genome include all sewage and biosolids samples (n= 13) from the United States and Europe (0.15% ± 0.09%), one out of two sewage samples from Africa (0.07% of reads), two out of six sewage samples from Asia (0.01% of reads from Shanghai and 0.05% of reads from Singapore), and one out of two bat guano samples (0.30% of reads). In addition, the reads from the United States and Europe sewage samples were mapped against other commonly investigated viral MST markers to compare the abundance of crAssphage to pepper mild mottle virus (PMMoV), human adenovirus F (HAdV), human polyomavirus BK (HPyV), Torque teno virus (TTV), and norovirus GII (NoV). Previous studies have investigated the utility of PMMoV [24, 114], HAdV [94, 95, 114], HPyV [94, 95, 114], TTV [114], and NoV [94, 95] as human MST markers, and PMMoV has been identified to be more abundant in sewage than other commonly used viral markers [24, 114]. The average number of reads mapped from U.S. and Europe metagenomes to each viral

genome is shown in the inset of Figure 3.1 (Data in Table A2). While PMMoV was more abundant than HAdV, HPyV, TTV, and NoV, crAssphage was the most abundant virus tested. All samples had a low percentage of reads mapped to PMMoV compared to crAssphage, with only 0.01% ± 0.01% of reads mapping to the PMMoV genome for samples from the US. The European sample (Barcelona) had a higher percentage of sequences mapping to the PMMoV genome (0.23%) than the crAssphage genome (0.15%), potentially reflecting the diet of the region. Overall, the percentage of reads mapped to the PMMoV genome (0.02% ± 0.06%) was significantly less ($p < 0.001$) than the percentage mapped to the crAssphage genome (0.15% ± 0.09%). This, together with metagenomic evidence for high crAssphage abundance in the gut, suggests that crAssphage may overcome detection limitations found with other viral markers.

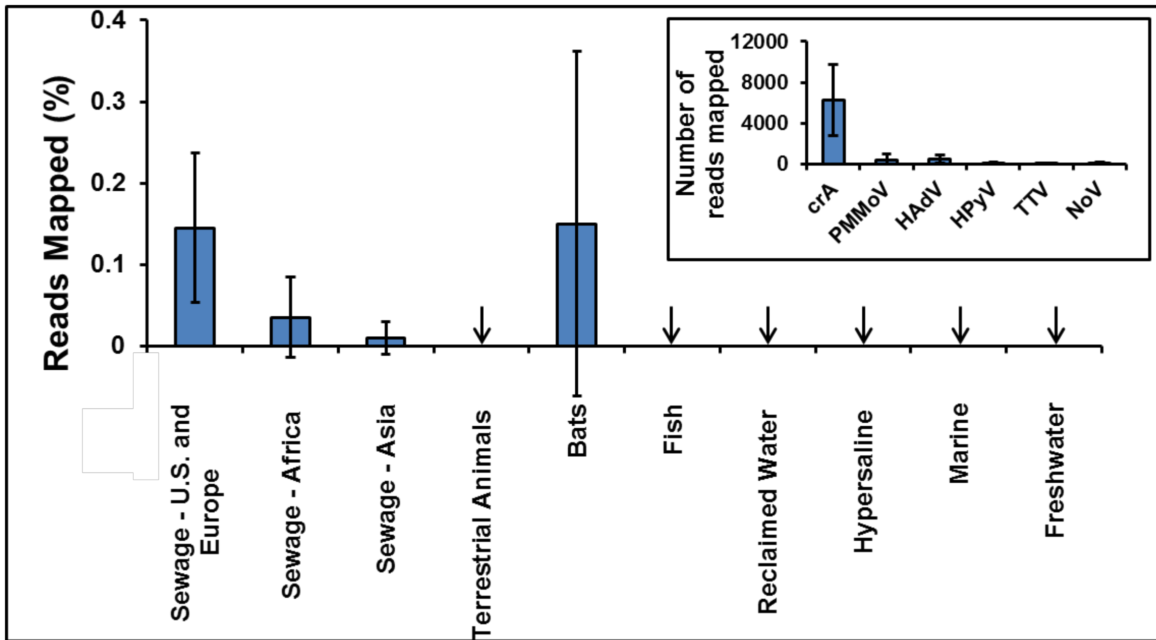

Figure 3.1: Percent of sequence reads mapped to the crAssphage genome for metagenomes from different sources. Error bars indicate ± 1 standard deviation. Downward arrows indicate that no sample in that category had at least 0.01% of reads mapped to the crAssphage genome. Sample sizes are indicated as either viral (V) or whole microbial metagenome (M): sewage-US and Europe (V_n=13), sewage-Africa (V_n=2), sewage-Asia (V_n=6), terrestrial animal (V_n=4, M_n=12) (cow=5, chicken=4, rodents=3, canine=2, pig=1, sea lion=1), bat (V_n=2), fish (V_n=2), reclaimed water (V_n=6), hypersaline water (V_n=12), marine water (V_n=9), freshwater (V_n = 6). Inset: Number of reads mapped for all sewage and biosolids samples from the US and Europe to various viral genomes (crA-crAssphage, PMMoV-pepper mild mottle virus, HAdV-human adenovirus, HPyV-human polyomavirus, TTV-Torque teno virus, NoV-norovirus).

CrAssphage genome coverage was calculated for all samples with at least 0.01% read mapping (Figure 3.2). Excluding samples from Addis Ababa and Singapore, which had a low total number of mapped reads, all sewage and anaerobic digester samples showed a high coverage of the crAssphage genome (68% ± 14%). The Shanghai sewage sample had only 0.01% of reads mapping to the genome; however, this sample exhibited the second highest coverage of the crAssphage genome (87%), likely due to the large nature of the dataset (over 35 million sequence reads). The bat guano sample also showed a relatively high fraction of coverage of the crAssphage genome (39%); however, not as high as would be predicted based upon the large number of reads that mapped to the genome.

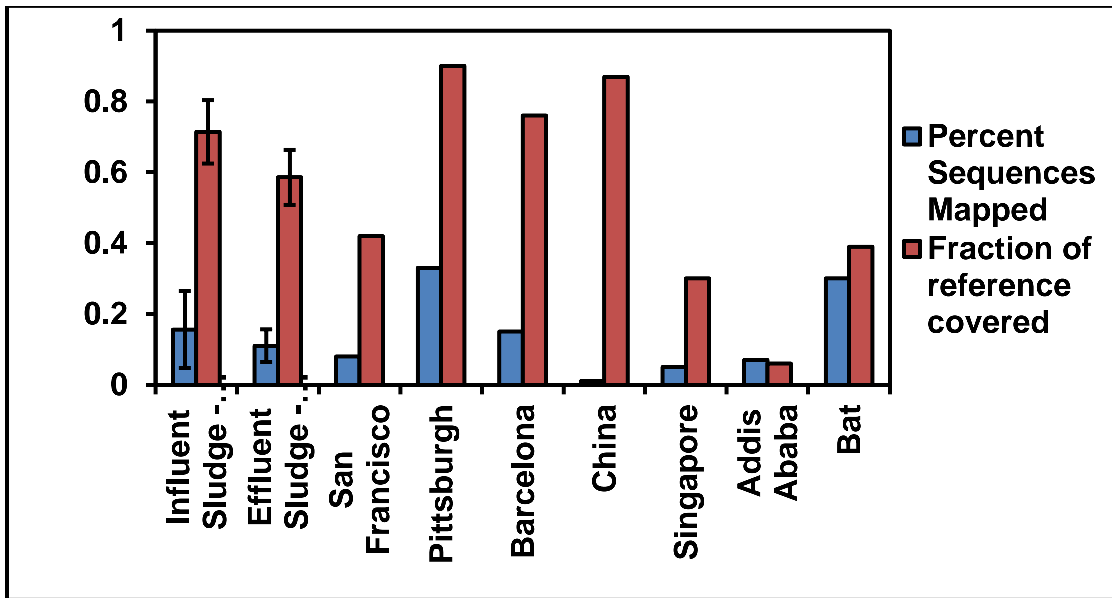

Figure 3.2: The percent of sequences mapped for individual samples compared to the fraction of the crAssphage genome covered by the mapped reads. Fraction of reference covered was only computed and displayed here if the percent of sequences mapped was at least 0.01%. Error bars for the influent and effluent sludge from the U.S. indicate ± 1 standard deviation.

Figure 3.3 shows the coverage of the crAssphage genome from read mappings for three samples: sewage from Pittsburgh, USA; sewage from Shanghai, China; and bat guano from China. For the bat guano sample, nearly half of the mapped reads (46%) mapped to orf00045 in the crAssphage genome. These reads mapped to one region of orf00045, representing only 36% coverage of orf00045. The median of the normalized percent of reads mapping to the crAssphage genome for bat guano, Pittsburgh sewage, and Shanghai sewage was 0.32, 0.84, and 0.85, respectively, indicating less even coverage of the crAssphage genome by the bat guano sample. This implies crAssphage may share some homology with a phage in the bat virome and care should be taken to eliminate this region in future assay development.

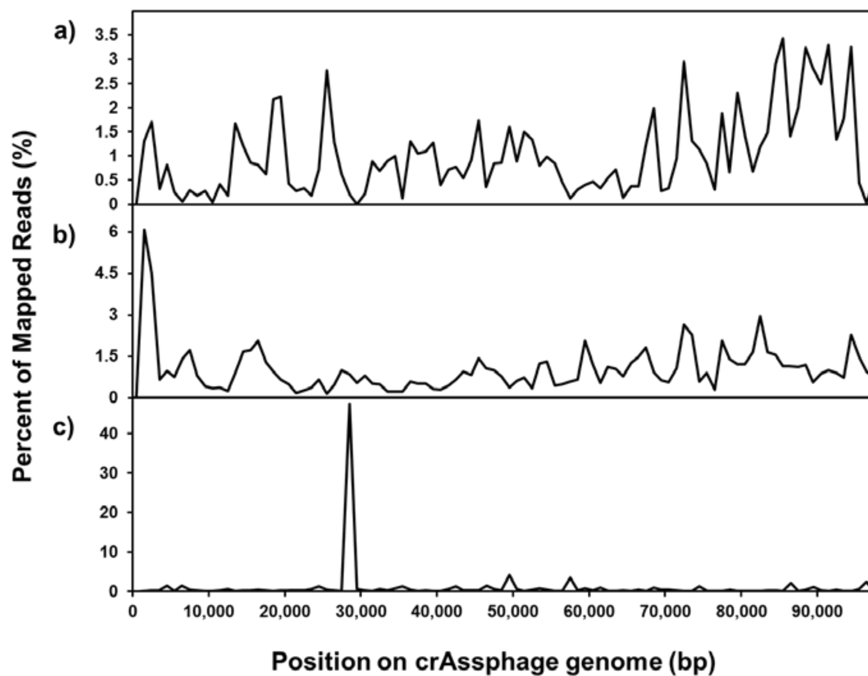

Figure 3.3: Percent of mapped reads versus position on the crAssphage genome for a) raw sewage from Pittsburgh, PA, USA. b) raw sewage from Shanghai, China c) bat guano from China

Manual inspection of sewage and bat guano read mapping identified several regions exhibiting promise for development as an MST marker. These regions were entered into PrimerBlast [152] and searched against the NCBI nr database to investigate whether these regions contain crAssphage specific markers. Regions that warrant further investigation include (approximate values of crAssphage genome bp positions): 1770-1870, 78100-78270, 83860-83970, 88370-88470, 90120-90280, and 93160-93340. Primers were found for all of these regions that were either specific only to crAssphage or to crAssphage and an uncultured organism clone from the human gut metagenome (accession GQ873945.1).

Within sewage samples, crAssphage exhibits a strong geographic dependence, being more abundant in the US and Europe than Africa or Asia, consistent with previously recognized geographic dependence on the microbial ecology of the gut microbiome [153]. For example, a recent study noted an immediate shift in gut microbial ecology for an individual when traveling to the developing world, which was reversed when the individual returned to the US [154]. The factors driving this shift remain unresolved, but may play a strong role in the abundance of crAssphage in sewage. Although crAssphage was less abundant, its presence in samples from both Africa and Asia suggest that it is present worldwide. However, more research is warranted to investigate the full geographic prevalence of crAssphage to determine if it would be a suitable marker in these areas.

The metagenomic approach applied here for identification of crAssphage for microbial source tracking has certain limitations. Samples for evaluation are constrained to metagenomes available in public databases. Additionally, the composition of samples is biased by the sampling and processing approaches used in each respective study while sample depth is limited to that obtained by the initial studies. Additional empirical data (e.g. qPCR) is necessary to confirm the

absolute abundance of crAssphage in human samples and absence in non-human samples. Despite these limitations, a metagenomic approach represents a considerable time and cost savings over laboratory methods for initial evaluation of MST markers. Further research is necessary before a crAssphage marker can be utilized in the field. Next steps include marker selection and testing in human and animal fecal samples as well as contaminated waters to assess the field sensitivity, specificity, limit of detection, geographic and seasonal variations, and co-occurrence of crAssphage with human pathogens. However, this approach represents a valuable first step in verifying that these next steps should be pursued.

This study sought to understand the utility of the newly discovered crAssphage as an MST marker for human fecal pollution by utilizing publicly available metagenomic data. Overall, this study has demonstrated that crAssphage is a strong candidate for a human MST marker based on its ubiquity in the human gut and human sewage and its absence in other environments, including other animals' guts. In addition, crAssphage has been found to be even more abundant than other proposed viral markers and viral pathogens, decreasing challenges associated with the dilution of marker abundance in natural waters that has been observed for other viral markers. Further, this study has demonstrated the utility of publicly available metagenomic data for initial evaluation of a new MST marker, a powerful approach that allows rapid results with limited economic and time investment.

4.0 QUANTITATIVE CRASSPHAGE PCR ASSAYS FOR HUMAN FECAL POLLUTION MEASUREMENT

This work has been published as:

Stachler, E.; Kelty, C.; Sivaganesan, M.; Bibby, K.; Shanks, O. Quantitative CrAssphage pcr assays for human fecal pollution measurement. *Environ. Sci. Technol.* 2017, 51(16), 9146-9154.

This work is also the subject of a patent application:

Stachler, E.; Bibby, K.; and Shanks, O. Cross-Assembly Phage DNA Sequences, Primers and Probes for PCR-based Identification of Human Fecal Pollution Sources. Application Number: 62/386,532

Environmental waters are monitored for fecal pollution to protect public health and water resources. Traditionally, general fecal indicator bacteria are used; however, they cannot distinguish human fecal waste from other animal pollution sources. Recently, a novel bacteriophage, crAssphage, was discovered by metagenomic data mining and reported to be abundant in and closely associated with human fecal waste. To confirm bioinformatic predictions, 384 primer sets were designed along the length of the crAssphage genome. Based upon initial screening, two novel crAssphage qPCR assays (CPQ_056 and CPQ_064) were designed and evaluated in reference fecal samples and water matrices. The assays exhibited high specificities (98.6%) when tested against an animal fecal reference library and crAssphage

genetic markers were highly abundant in raw sewage and sewage impacted water samples. In addition, CPQ_056 and CPQ_064 performance was compared to HF183/BacR287 and HumM2 assays in paired experiments. Findings confirm viral crAssphage qPCR assays perform at a similar level to well established bacterial human-associated fecal source identification approaches. These new viral based assays could become important water quality management and research tools.

4.1 INTRODUCTION

Many environmental waters are polluted with human fecal waste originating from numerous sources such as leaking sewer lines, faulty septic systems, improperly connected downspouts, and combined sewer overflows. Human fecal waste can harbor disease-causing pathogens that contribute to poor public health, reduced ecological outcomes, and economic burdens. Many public health managers rely on general fecal indicator methods (e.g. *E. coli* and enterococci) to monitor fecal pollution levels, which do not discriminate between human and other potential animal sources of fecal pollution. General indicators provide limited information which prevents focused remediation, because many areas are polluted by a combination of human, agricultural, and/or wildlife sources. Information on human waste (i.e. sewage) is particularly important because it may pose a greater risk to public health compared to fecal pollution from other animal sources [18, 19, 52]. To compliment general indicator measurements and better characterize human fecal pollution, many researchers and water quality managers use fecal source identification technologies.

Most currently available human fecal source identification technologies target fecal bacteria, mainly *Bacteroides* species [55]. These fecal bacteria are abundant in the human gut and sewage, facilitating their detection in diluted environmental water samples. Bacterial human fecal genetic markers, such as HF183/BacR287 and HumM2 are highly human-associated and reproducible in multiple laboratory validation studies [17, 55, 61]. Although bacterial methods exist for human fecal source characterization, technologies targeting viruses are also needed [8, 87, 155].

Enteric viruses, such as norovirus, adenovirus, and enterovirus are reported to be the dominant etiological agents of waterborne and shellfish-borne disease [1, 2]. Several studies suggest enteric viruses react to environmental and waste treatment conditions in markedly different ways compared to bacterial fecal indicators [87, 156-159]. In addition, waterborne viral outbreaks have occurred when general bacterial fecal indicators are not detected or are below regulated levels [1]. Reliance solely on bacterial indicators limit the ability of water quality managers to link measures of human fecal pollution with public health risk; thus, viral human-associated technologies offer an attractive alternative to bacterial fecal source identification methods. Researchers have previously recognized the potential value of viral human-associated methodologies, leading to the development of technologies targeting enterovirus [99-101, 160, 161], adenovirus [20, 21], norovirus [22, 93], polyomavirus BK and JC [23], somatic coliphage [162, 163], *Bacteroides* phages [125-127], and pepper mild mottle virus [24], among others. However, a recent multiple laboratory study evaluated the performance of many of these virus methods and concluded that the technologies tested either lacked sensitivity or exhibited poor specificity, potentially limiting suitability for widespread water quality management applications

[94]. A recent comparison of human polyomavirus levels to predicted public health risk also illustrates that more sensitive viral methodologies are needed [107].

An ideal viral human-associated method for environmental water quality testing would target a virus that is both highly human-associated and consistently abundant in human waste sources. Recently, a novel bacteriophage, "crAssphage", was described via metagenome cross-assembly and was predicted to be a *Bacteroides* phage by co-occurrence profiling [32]. The double-stranded DNA crAssphage putative genome was assembled from shotgun metagenomic libraries isolated from an individual human fecal sample [32]. Further bioinformatic testing predicted that the crAssphage genome is highly abundant and was identified in 73% of human fecal metagenomes surveyed [32]. A subsequent metagenome survey detected crAssphage in sewage from the United States and Europe, while crAssphage was absent in other environments, such as non-human fecal samples and water environments [164]. In addition, the crAssphage genome is estimated to be up to 10 times more abundant in sewage than other known human-associated viruses, including norovirus and adenovirus [164].

Near-ubiquity across human fecal metagenomes and the high abundance compared to other sewage-derived viruses, combined with potential human specificity, motivates the development of crAssphage as a fecal source identification technology [164]. However, several unknown issues remain that must be addressed for the successful development of a crAssphage fecal source identification tool. For instance, the crAssphage genome likely represents a viral quasi-species consensus sequence compiled from a collection of DNA regions with unknown sequence variability. Furthermore, most information about the crAssphage genome has been generated from computer predictions with minimal laboratory testing to verify findings. Extensive laboratory testing of fecal samples gathered from a wide variety of animal species and

sewage collected across a broad geographic range is necessary to evaluate the suitability of the crAssphage genome for human fecal source identification applications.

The goals of the present study are to survey the crAssphage genome for human-associated genetic regions, develop qPCR methods as potential future environmental water quality monitoring tools, and compare their performance to top performing bacterial human-associated technologies. We employed a "biased genome shotgun strategy" where select genetic regions of the crAssphage genome were screened using end-point PCR for highly specific and abundant human-associated fecal pollution genetic regions. These genetic regions were subsequently utilized to develop two novel qPCR fecal source identification assays. Findings suggest that high throughput laboratory screening of novel virus genomes discovered through metagenomic DNA sequence mining is a successful strategy to develop host-associated qPCR methods that may be important for future research and water quality management activities.

4.2 MATERIALS AND METHODS

4.2.1 Sample Collection

Individual fecal samples (n=222) were collected from various locations across the continental United States as previously described [61]. Animal fecal samples represent ten species including *Anser* spp. (Canada goose; n=18), *Canis familiaris* (dog, n=41), *Bos taurus* (cow, n=61), *Larus* spp. (gull, n=25), *Equus caballus* (horse, n=20), *Cervus canadensis* (elk, n=20), *Gallus gallus* (chicken, n=11), *Sus* scrofa. (pig, n=9), *Castor canadensis* (beaver, n=8), and *Odocoileus virginianus* (deer, n=9) (see Table B1 for sample details). Each fecal sample was collected from

a different individual as previously described [61] and stored at -80°C until time of DNA extraction (<18 months). Primary influent sewage samples were collected at nine geographically distributed wastewater treatment plants within the United States (Table B2) as previously described [61]. Briefly, one liter of primary influent was collected and immediately packed in ice and shipped overnight to Cincinnati, OH USA for laboratory testing. DNA extraction of sewage samples was performed within 48 hrs of collection as described below. Finally, as a proof of concept pilot study, surface water samples were collected from the Heiserman Stream (East Fork Watershed, southwest OH USA) in close proximity to a treated sewage discharge area. These samples were collected in a sterile 1-L container, immediately stored on ice and transported to the laboratory for DNA extraction and testing (<4 hrs).

4.2.2 DNA Extraction and Quantification

DNA was extracted from 10 mL of primary influent sewage with the QIAamp Blood Maxi Kit according to manufacturer's instructions, except Buffer AVL was substituted for Buffer AL (Qiagen, Valencia, CA). DNA was extracted from individual animal fecal samples using the DNA-EZ Kit (GeneRite, North Brunswick, NJ USA) substituting Buffer AE (Qiagen, Valencia, CA USA) for the elution buffer in a modified protocol of the manufacturer's instructions. Briefly, fecal slurries were made by adding molecular grade PBS and fecal matter to the bead mill tubes and were homogenized in a bead beater at 6 m/s for 30 s. After a prolonged centrifugation, 760 µL Binding Buffer was added to recovered supernatant and the manufacturer's instructions were followed eluting with molecular grade water warmed to 60°C. For environmental water samples, a 200 mL sample was concentrated to a final volume of approximately 150 µL using an automated Concentrating Pipette with a single-use ultra-filtration

hollow fiber polyethersulfone tip following manufacturer's instructions (InnovaPREP®, Drexel, MO USA). DNA extraction of concentrate was performed with the DNA-EZ Kit (GeneRite, North Brunswick, NJ USA) as described above for fecal sample processing. Water sample DNA extracts were stored at 4°C in GeneMate Slick low-adhesion microcentrifuge tubes (ISC BioExperess, Kaysville, UT USA) until time of amplification (< 24 h). Fecal and primary influent sewage DNA extract concentrations were determined with a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE USA), diluted to 0.5 ng/µL to normalize sample test quantities for performance testing, and stored in low-adhesion microcentrifuge tubes at -20°C (<6 months). For each batch of DNA extractions, three method extraction blanks with purified water substituted for fecal, sewage, or environmental water were performed to monitor for potential contamination.

4.2.3 Selection of Candidate Genetic Regions for PCR-Based Assay Development

To identify candidate genetic regions for the development of human-associated fecal source identification methods, select portions of the putative ~97 kbp crAssphage genome (accession: JQ995537) [32] were identified for end-point PCR laboratory testing. Due to the reported potential for rapid DNA mutation rates in the human gut virome, such as those described for *Microviridae* [165], efforts were focused on predicted coding regions to select for sequences with some evidence of genetic conservation. Metaviromic islands were also excluded due to increased genetic diversity leading to under-recruitment in metaviromes suggesting the potential for low abundance in environmental samples [32, 166]. In addition, regions bordering modular junctions were eliminated. Finally, since orf00045 has homology with bat guano virome sequences [164], it was not considered for human-associated crAssphage assay development.

4.2.4 Candidate Primer Set Design

A total of 384 end-point PCR primer pairs were designed to amplify selected crAssphage genomic regions. Primer pairs were designed and tested *in silico* using Primer-BLAST [152] with default parameters, except PCR product length was constrained to 90-180 bps and primer pair specificity was evaluated using the nr database (May, 2015). Only primer pairs that generated BLAST hits (E-value < 30000) to crAssphage or clone DNA sequences from human gut metagenome projects were selected as candidate primer pairs. Genetic regions where no primer sets met design criteria were eliminated from further consideration.

4.2.5 End-point PCR Amplifications

Each 25 μL end-point PCR amplification consisted of TaKaRa *Ex Taq* Hot Start PCR reagents (Clontech Laboratories, USA), 100 nM each forward and reverse primer, 0.8 μg bovine serum albumin (BSA; Sigma-Aldrich, St. Louis, MO USA), 2 μL template DNA, and molecular grade water. End-point PCR tests were performed in duplicate or in triplicate on a Tetrad 2 Thermocycler (Bio-Rad Laboratories) under the following conditions: 94°C for 5 mins followed by 40 cycles of 94°C for 40 s, 57°C for 1 minute, and 72°C for 30 s, followed by a final extension at 72°C for 10 minutes. To monitor for potential sources of extraneous DNA during end-point PCR amplification, a minimum of two no-template controls (reactions contained additional purified water instead of template DNA) were performed with each instrument run. PCR products were verified on a 2% agarose gel with 1% lithium borate and 1X GelStar (Lonza, Rockland, ME USA) and visualized on a Gel Logic 100 Imaging System (Eastman Kodak Company, Rochester, NY USA).

4.2.6 Candidate End-Point PCR Primer Set Evaluation

In order to determine which candidate crAssphage genetic regions have human fecal source identification potential, all 384 primer sets were tested with end-point PCR in a three-round process. In the first round, candidate primer sets were challenged against two fecal DNA composites: sewage and non-human. A sewage DNA composite was created by combining primary influent sewage DNA from three geographic locales (1 ng total DNA/reaction, 0.33 ng DNA/reaction from each sample) and was used to identify the presence or absence of candidate genetic regions in a known human fecal pollution source. A non-human DNA composite was created from cow (n=9), dog (n=9), goose (n=9), and pig (n=9) fecal DNA (4 ng total DNA/reaction, 1 ng DNA/reaction from each animal group) and was used to determine the presence of candidate genetic regions in non-target fecal pollution sources. All primers were tested in duplicate against DNA composites. Candidate primer sets proceeded to a second round of testing if the following criteria were met: (1) a PCR product of expected size was present when primary influent sewage DNA composite was used as template, (2) PCR product of expected size was absent when the non-human DNA composite was used as a template, (3) low amplification efficiency was not observed in reactions with sewage DNA composite as the template as evaluated by manual inspection, and (3) absence of any spurious PCR products noticeably different in size from the expected PCR product, including primer dimers.

In round two, remaining candidate primer sets were challenged against diluted preparations of the primary influent sewage DNA composite (0.1 ng/reaction, $1x10^{-2}$ ng/reaction, and $1x10^{-3}$ ng/reaction) and a higher concentration of individual animal group composites for cow (n=9), dog (n=9), goose (n=9), and pig (n=9) using 5 ng of total DNA per reaction. Candidate primer sets proceeded to a third round of testing under the following conditions: (1)

amplification of expected size in triplicate reactions when 1×10^{-2} ng total DNA per reaction from primary influent sewage composite was used as template, (2) absence of expected PCR product size in all reactions when a non-human DNA composite was used as template, and (3) absence of spurious PCR byproducts, including primer dimers.

Round three represented the most rigorous performance screening step for candidate end-point PCR primer sets. Testing began with specificity determination from an expanded fecal reference collection (n=70 individual samples), followed by geographic distribution characterization in primary influent sewage samples collected from nine different locations, and ending with limit of detection (LOD) assessment. Reference fecal samples for specificity screening included cow (n=9), goose (n=8), dog (n=9), pig (n=9), horse (n=9), elk (n=9), deer (n=9), and beaver (n=8). All candidate primer sets passing round two were challenged with individual DNA preparations at 1 ng of total DNA per reaction. Specificity was defined as the proportion of non-human samples testing negative for a crAssphage genetic region. Only candidate primer sets with an observed specificity of 100% proceeded to geographic distribution testing. Sewage distribution characterization entailed testing of 1 ng of total DNA per reaction isolated from nine primary influent sewage samples collected from different locations across the continental United States (Table B2). Candidate primer sets with ≥ 95% detection frequency were eligible for LOD assessment. LOD_{95} was measured based on repeated testing (40 replicates per primer set) of primary influent sewage composite serial dilutions (10, 1, 0.1, 1×10^{-2}, and 1×10^{-3} ng total DNA per reaction) consisting of equal DNA mass from samples collected from all nine geographic locales. LOD_{95} was defined as the lowest dilution concentration where a minimum of 95% (38 of 40) of reactions yielded an amplification product of the expected size.

4.2.7 DNA Sequence Verification of Top Performing End-Point PCR Primer Sets

To verify that candidate primer sets passing round three screening were amplifying the intended crAssphage genetic region, amplification products from the round one primary influent sewage composite and an environmental water sample with known human sewage pollution impairment (Heiserman Stream, OH USA) were sequenced and evaluated. PCR was performed using primer sets passing round three, using the same amplification conditions as above. PCR products were cloned into plasmid vector pCR2.1-TOPO and transformed into TOP10 chemically competent cells using the TOPO TA Cloning Kit as described by the manufacturer (Invitrogen, Thermo Fischer Scientific, USA). Transformed *E. coli* colonies plated on LB plates with kanamycin and X-gal for blue/white screening were sent to GENEWIZ for sequencing (South Plainfield, NJ USA). Sanger sequencing was performed from transformed bacterial colonies for each primer-template combination using the M13R primer for amplification. PCR products were aligned with the previously reported crAssphage sequence (accession: JQ995537) [32] using CLC Genomics Workbench 8.5.1 (Qiagen, Valencia, CA USA).

4.2.8 CrAssphage qPCR Assay Development

Candidate primer sets passing round three end-point PCR testing were adapted to TaqMan® qPCR chemistry. Primers and probes for putative human-associated crAssphage genetic regions were designed using default parameters of the Primer Express version 3.0.1 software (Thermo Fisher Scientific, USA). Fluorogenic minor binding groove (MGB) probes were 5' labeled with 6-carboxyfluorescein.

4.2.9 qPCR Amplifications

Five qPCR assays were used in this study: two novel crAssphage assays (this study) and two previously reported human-associated bacterial fecal source identification methods (HF183/BacR287 and HumM2), as well as an environmental water sample processing control assay (Sketa22) [16, 17, 60]. Each 25 µL qPCR reaction was composed of 1X TaqMan® Environmental Master Mix 2.0 (Thermo Fisher Scientific, USA), 5 µg BSA, 1 µM of each primer, 80 nM 6-carboxyfluorescein (FAM)-labeled probe, 80 nM VIC-labeled probe (HF183/BacR287 and HumM2 only), 2 µL template DNA, and molecular grade water. All qPCR tests were performed in triplicate using the QuantStudio™ 6 Flex Real-Time PCR System (Thermo Fisher Scientific, USA). The thermal cycling profile for all assays was 10 minutes at 95°C followed by 40 cycles of 15 s at 95°C and one minute at 60°C. The threshold for qPCR assays was manually set to 0.03 (crAssphage, HF183/BacR287, and Sketa22) or 0.08 (HumM2) and quantification cycle (C_q) values were exported to Microsoft Excel. Six y-intercept control reactions (standard reference material at 100 copies/reaction) were performed with each instrument run to utilize a mixed calibration model approach [167]. To monitor for potential contamination, six no-template controls were performed with each instrument run. Amplification inhibition was monitored in all DNA extracts using the HF183/BacR287 and HumM2 IAC procedures as previously reported [168].

4.2.10 qPCR Standard DNA Material Preparation

Standard DNA material consisted of a customized gBlock™ gene fragment containing target sequences for crAssphage, HF183/BacR287, and HumM2 standard curve generation and an

internal amplification control (IAC) plasmid construct for HF183/BacR287 and HumM2 amplification inhibition screening (Integrated DNA Technologies, Coralville, IA USA) [168]. Standard DNA concentrations were determined with a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE USA). For standard curve reference material, five dilutions were prepared to contain 10 to 1×10^5 copies/2 µL. IAC reference DNA material was prepared as previously described [168]. All reference DNA material preparations were stored in GeneMate Slick low-adhesion microcentrifuge tubes (ISC BioExpress, Kaysville, UT USA) at -20°C prior to use (<3 months).

4.2.11 Performance Testing of crAssphage qPCR Assays

To investigate the suitability of newly developed technologies for human fecal source identification application, the performance of crAssphage qPCR methods was evaluated in a series of head-to-head experiments with HF183/BacR287 and HumM2 bacterial human-associated methods [16, 17]. Calibration model performance including amplification efficiency ($E = 10^{(-1/slope)}-1$), lower limit of quantification (LLOQ), and precision (% coefficient of variation) at 10 copies per reaction were calculated from six standard curves generated from independent instrument runs. LLOQ (log_{10} copies/reaction) was defined as the upper bound of the 95% credible interval from repeated measures of the 10 copy per reaction standard curve dilutions. Next, the abundance of each genetic marker was measured in primary effluent sewage samples (n=9) at a test concentration of 1 ng of total DNA per reaction (Table B2). The prevalence of putative human-associated genetic markers in non-human pollution sources was evaluated with a reference fecal collection consisting of 222 individual samples from 10 different animals (Table B1; test quantity = 1 ng total DNA/reaction). Prevalence was expressed both

quantitatively (log_{10} copies/ng of total DNA) and qualitatively (specificity = TNC/(TNC+TPI), where TNC represents the total number of negative individual samples that tested negative correctly, and TPI is the total number of individual samples that tested positive incorrectly). Finally, as a proof-of-concept pilot demonstration, genetic marker concentrations were estimated from two environmental water samples known to be impacted by human sewage pollution (Heiserman Stream, OH). Average log_{10} copies per ng of total DNA (sewage) and log_{10} copies per reaction (water) with 95% credible intervals were determined (mean Cq for each sample group and assay combination transformed using respective mixed calibration model followed by a nested analysis of variance to estimate standard deviation values) and compared to identify similarities and differences between qPCR genetic marker concentrations.

4.2.12 Data Analysis

Mixed model calibration models, unknown DNA concentration estimates, and credible intervals were determined using a Monte Carlo Markov Chain (MCMC) approach [167]. MCMC calculations were performed using the publically available software WinBUGS, version 1.4.1 (http://www.mrc-bsu.cam.ac.uk/bugs).

4.3 RESULTS

4.3.1 Putative Human-Associated crAssphage Genetic Regions and Candidate Primer Set Design

A total of 46,564 bp (48%) of the crAssphage genome were selected for end-point PCR screening to identify potential human-associated genetic regions. Select genetic regions were excluded due to: (1) non-coding regions (8%), (2) metaviromic island motifs (32%), (3) modular junction regions (3%), (4) evidence of similarity with non-human or non-crAssphage sequences (3%), and (5) regions not amenable for PCR testing based on primer design parameters (e.g., product size and melting temperature restrictions (6%)) (Figure 4.1). In total, 384 end-point PCR primer sets were designed with 90% coverage (41,794 bp) of select putative human-associated genetic regions (Figure 4.1, Table B3).

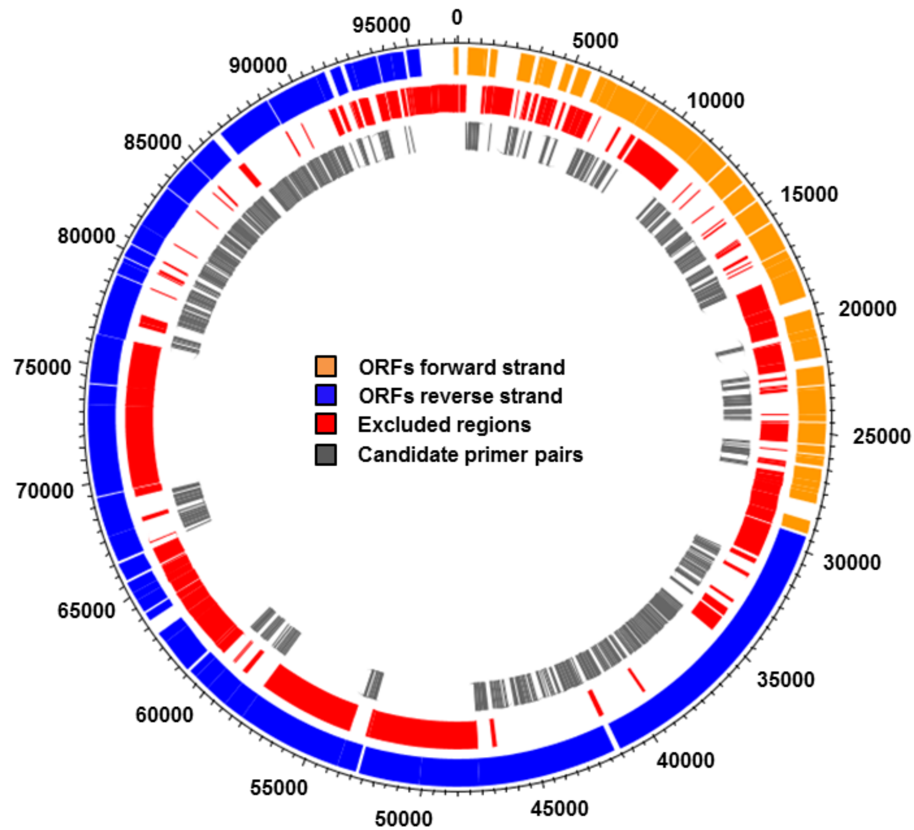

Figure 4.1: Map representation of the crAssphage genome. The outermost track represents the open reading frames (ORFs) on the forward and reverse strand of the crAssphage genome. The middle track represents the areas of the crAssphage genome that were eliminated from primer design, including non-coding regions, metaviromic islands, modular junction areas, non-target sequence homology, and regions unsuitable for primer design. The innermost track represents the location of the 384 end-point primer pairs designed in this study and their amplification products.

4.3.2 Identification of Human-Associated crAssphage Genetic Regions with End-Point PCR

Candidate primer sets were subjected to three rounds of performance testing to identify human-associated crAssphage genetic regions. The first round of testing comprised of testing the primers against a sewage DNA composite and non-human animal fecal DNA composite. Round one testing eliminated 327 (85.2%) candidate primer sets: 34.6% (n=133) failing to yield a clearly distinct PCR product of the expected size in the sewage composite, 44.0% (n=169) generating spurious PCR products (including primer dimerization byproducts), and 4.7% (n=18) yielding false-positive results in non-human composite tests. Forty-nine (12.8%) candidate primer sets failed more than one first-round criteria. A total of 57 candidate primer sets were deemed eligible for round two testing.

In round two testing, primer sets were challenged against lower dilutions of sewage composite DNA and higher concentrations of non-target animal fecal composite DNA. A total of 51 primer sets were eliminated due to: amplification product of expected size when non-human samples were used as DNA template (n=24), failure to consistently yield a PCR product of the expected size when $1x10^{-2}$ ng total sewage composite DNA per reaction was used as template (n=14), and/or evidence of spurious PCR byproducts, including primer dimerization (n=40). Twenty-seven primer sets failed more than one criteria. False positives observed with each non-human animal group tested were: pig (n=15), cow (n=15), canine (n=6), and goose (n=1). Six primer sets passed round two testing, including crAss028, crAss056, crAss064, crAss301, crAss303, and crAss375.

Round three testing included specificity determination with an expanded reference fecal collection, characterization of geographic distribution in sewage, and a limit of detection

(LOD_{95}) assessment (Table B4). Primer sets crAss056 and crAss064 exhibited the best performance with 100% specificity and 100% detection in geographic sewage samples and were subject to LOD_{95} assessment. Both primer sets yielded an LOD_{95} of $1x10^{-2}$ ng total DNA per reaction. Primer sets crAss064 and crAss056 were detected in 52.5% and 45% of test replicates, respectively, at a DNA template concentration of $1x10^{-3}$ ng per reaction.

4.3.3 DNA Sequencing Verification

End-point PCR products from crAss056 and crAss064 primer sets were sequenced from a primary influent sewage composite and human fecal pollution impacted environmental water sample to confirm amplification of the expected crAssphage sequences. Sequencing efforts resulted in 91 sequences (Figure B1). Alignment of crAss056 sequences indicated that 84.1% (37 of 44) of sequences exhibited 100% similarity to the corresponding reference crAssphage genome region (accession: JQ995537; 14735 to 14836 bp). Five additional variants designated B, C, D, E, and F were observed with 1 mismatch each (99% similarity to crAssphage genetic region). Primer set crAss064 alignments yielded 74.5% (35 of 47) of sequences with 100% similarity to reported crAssphage genomic region (16058 to 16152 bp). Variant D was observed in 12.8% (6 of 47) of the sequences with a one base pair substitution. The remaining six crAss064 sequences each exhibited sequence similarities ranging from 98% (2 mismatches) to 99% (1 mismatch) designated variants B, C, E, F, G, and H.

4.3.4 Performance of crAssphage qPCR Assays

Candidate primer sets crAss056 and crAss064 were adapted as CPQ_056 and CPQ_064 to TaqMan® qPCR chemistry, respectively (sequences in Table 4.1). A series of paired experiments were performed to characterize new crAssphage qPCR assays with established HF183/BacR287 and HumM2 methods. Calibration model performance metrics are reported in Table B5 and include slope, y-intercept range, amplification efficiency (*E*), LLOQ range, and precision at 10^1 copies per reaction. All assays had a range of quantification from 10^1-10^5 copies per reaction (full range of tested standard concentrations). CPQ_056 and CPQ_064 both exhibited a specificity of 98.6% cross-reacting with the same three individual samples from gull (n=2) and dog (n=1), while HF183/BacR287 (100%) and HumM2 (99.5%; elk=1) yielded slightly higher performance levels. CPQ_056 and CPQ_064 target $\log_{10}$ copies per ng of total DNA concentrations were ≤1.33±0.04 in the two cross-reacting gull samples and ≤ 2.60±0.01 in one dog sample. HumM2 was detected in a single elk sample (1.02±0.06 $\log_{10}$ copies per ng of total DNA). Genetic marker $\log_{10}$ copies per ng of total DNA concentrations in primary influent sewage samples collected from different geographic locations ranged from 1.49±0.05 to 3.37±0.05 (CPQ_056), 1.83±0.04 to 3.47±0.05 (CPQ_064), 1.55±0.02 to 3.18±0.02 (HF183/BacR287), and 1.13±0.02 to 2.09±0.02 (HumM2). Total $\log_{10}$ copies per reaction concentrations in polluted environmental water samples ranged from 2.12±0.04 to 2.50±0.04 (CPQ_056), 2.33±0.03 to 2.55±0.03 (CPQ_064), 2.28±0.07 to 2.45±0.07 (HF183/BacR287), and 1.06±0.06 to 1.49±0.06 (HumM2). Wastewater qPCR reactions contained 1 ng of template DNA extracted from 10 mL of wastewater while environmental water qPCR reactions contained 2 µL of DNA extracted from a total volume of 200 mL of impacted water. A comparison of mean estimates with 95% credible intervals both indicated that primary influent sewage (Figure 4.2,

Panel A) and environmental water samples (Figure 4.2, Panel B) show no significant difference between CPQ_056, CPQ_064, and HF183/BacR287 results, while HumM2 measurements were significantly lower ($p<0.05$).

Table 4.1: CrAssphage qPCR assay oligonucleotides and targeted genomic regions

qPCR Assay	Primer/ Probe	Sequence 5’ → 3’	Genomic Region
CPQ_056	056F1	CAGAAGTACAAACTCCTAAAAAACGTAGAG	14731-14856
	056R1	GATGACCAATAAACAAGCCATTAGC	
	056P1	[FAM] AATAACGATTTACGTGATGTAAC [MGB]	
CPQ_064	064F1	TGTATAGATGCTGCTGCAACTGTACTC	16030-16177
	064R1	CGTTGTTTTCATCTTTATCTTGTCCAT	
	064P1	[FAM] CTGAAATTGTTCATAAGCAA [MGB]	

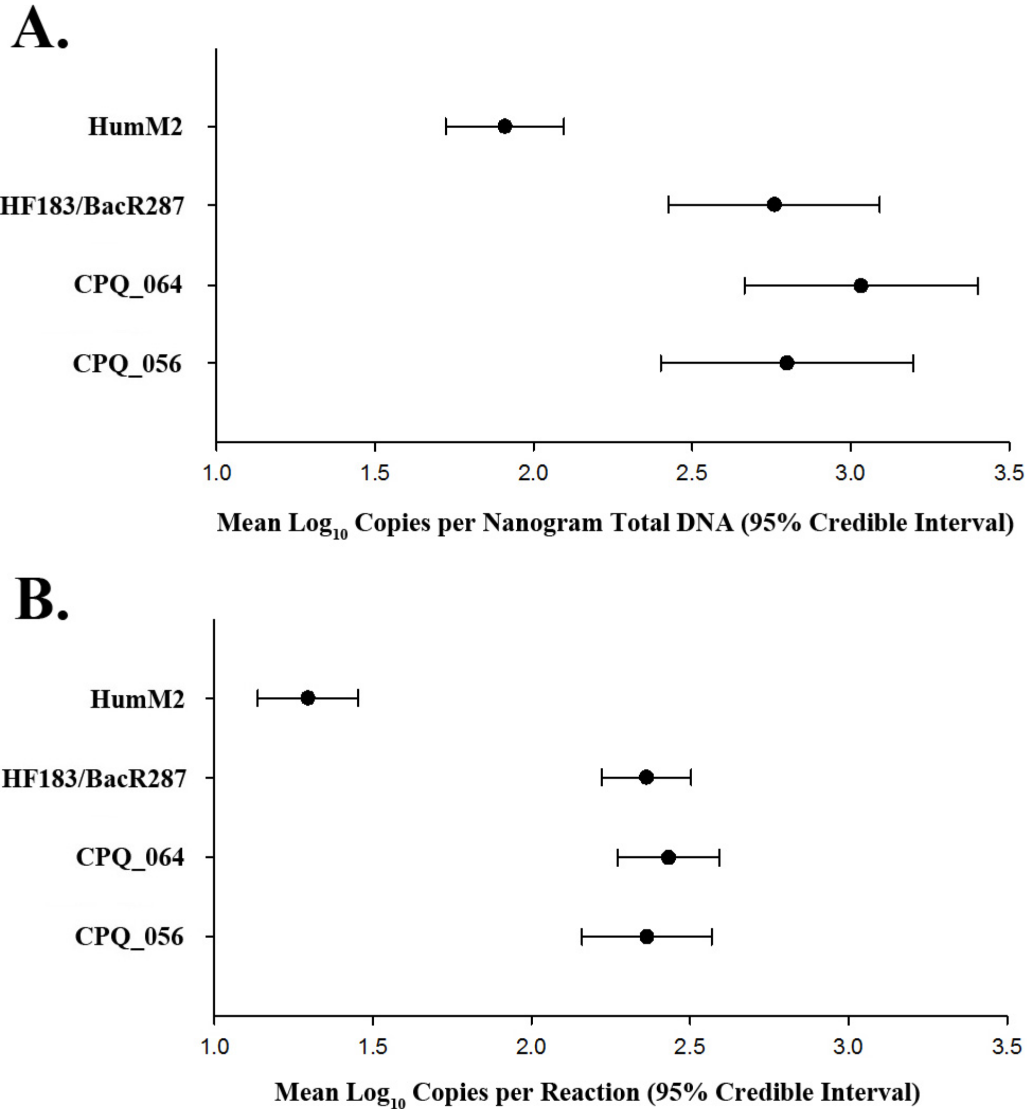

Figure 4.2: Abundance of crAssphage and bacterial human-associated qPCR targets in primary influent sewage (Panel A) and environmental water samples (Panel B). Values are reported as mean $\log_{10}$ copies estimates per ng of total DNA (Panel A) or per reaction (Panel B) with 95% credible intervals. Sewage qPCR reactions contained 1 ng of template DNA extracted from 10 mL of wastewater while environmental water qPCR reactions contained 2 µL of DNA extracted from a total volume of 200 mL of impacted water.

4.3.5 Experiment Controls

No template control amplifications indicated the absence of contamination in 99.7% of control reactions (n=1884). All method extraction blanks were negative ensuring no contamination was introduced during sample DNA extraction procedures. All DNA preparations exhibited no evidence of amplification inhibition except three fecal sample preparations, which were discarded from the study (data not shown). All environmental water samples showed no evidence of matrix interference as determined using the Sketa22 approach (data not shown).

4.4 DISCUSSION

4.4.1 Identification of Human-Associated CrAssphage Genetic Regions

Data mining of human fecal metagenomic DNA libraries recently identified a putative crAssphage genome [32]. Previously reported comparative sequence analyses suggest this genome is both highly abundant and broadly distributed in human fecal and sewage metagenomic DNA sequence libraries [32, 164]. To investigate the potential use of crAssphage for human fecal source identification qPCR method development, we interrogated 43% of the crAssphage genome via laboratory testing to identify candidate qPCR target genetic regions. Approximately 65% of end-point PCR candidate primer sets generated expected PCR products in primary influent sewage samples supporting bioinformatic predictions that the crAssphage genome is widespread in United States wastewaters [164]. In contrast, 35% of candidate primer sets targeting genetic regions selected by bioinformatic analysis were not detected in sewage. No

PCR product amplification from these primer sets may have been due to a lack of assay optimization (e.g., reaction mixture, thermal cycling conditions) or sequence variation at primer hybridization sites. Another plausible explanation could be low genetic conservation between individuals leading to reduced template availability in sewage DNA extracts. The crAssphage genome is reported as a consensus sequence of a quasispecies population isolated from a single fecal sample [32], whereas sewage is typically a mixture of fecal material contributed typically from thousands of individuals. Diversity between individuals in the crAssphage genome is limited. A recent study identified that patients in China were missing one open reading frame (ORF) and had low identity to another ORF of the crAssphage genome in their samples; however, these ORFs were identified as metaviromic islands and were not interrogated for primer design within this study [32, 169]. Additional studies are warranted to characterize within and between individual sequence variation in these genetic regions.

Nearly 95% (n=366) of end-point PCR primer sets in round one testing did not yield amplification products of the expected size with non-human animal sources used as DNA template. These findings support bioinformatic predictions of a close association of the crAssphage genome with human fecal waste [32, 164]. The bioinformatic analysis limited the number of false-positives observed in laboratory testing; however, some false-positives were identified, suggesting that parts of the crAssphage genome share homology with gut microbiome associated microorganisms of animals tested. Fourteen of the 18 false-positive results in round one testing were located within ORFs with reported homology to known or hypothetical proteins [32]. In addition, 169 (44%) of the primer sets yielded spurious PCR products (e.g., incorrect size, primer dimerization) during round one end-point PCR screening of sewage and fecal DNA preparations. Primer dimerization products can occur when a short region of complimentary

bases is shared between oligonucleotides in the same reaction while byproducts of the incorrect size may result from the improper annealing of primers or the potential amplification of pseudogenes or conserved sequence motifs in DNA template preparations. Since the generation of spurious PCR products leads to competition for reagents between the DNA target of interest and amplification byproducts, these genetic regions were not considered for qPCR method development. However, since the current study did not employ any optimization of PCR reaction conditions, these genome regions may be human-associated and may yield adequate methods with additional optimization.

After three rounds of screening, the crAss056 and crAss064 primer sets were selected for qPCR method development. These genetic regions represent the most human-associated and abundant candidate DNA targets based on end-point PCR amplification conditions and reference fecal and sewage collections utilized in this study. Both primer sets target the forward strand of the crAssphage genome (Figure 4.1). Primer set crAss056 (14712-14860 bp) amplifies a region within orf00024, which currently has no known protein homolog [32]. Primer set crAss064 (16038-16177 bp) targets a region within orf00025, which was previously reported to have homology with a DNA primase/helicase protein from *Veillonella* sp. [32], a bacterial genera commonly found in the intestines and oral mucosa of mammals. DNA sequencing efforts verified that crAssphage amplification products from primer sets crAss056 and crAss064 are conserved in primary influent sewage and environmental samples tested in this study (Figure B1), implying a level of genetic stability for the crAss056 and crAss064 target regions. Limited information exists on DNA mutation rates of intestinal viruses; however, several studies report considerable variation in the human gut virome in samples taken from different individuals [165, 170, 171]. In this study, we attempted to avoid genetic regions with high mutagenic or

recombination potential by focusing only on predicted protein coding regions without metaviromic islands or proximity to modular junction regions. The crAssphage assays may need to be monitored in the future to ensure DNA sequence stability of the targeted gene sequences.

4.4.2 Performance of crAssphage qPCR Assays

Systematic testing of 384 candidate primer sets identified two genetic regions (primer sets crAss056 and crAss064) that were selected for qPCR method development based on the study design. The performance of crAssphage CPQ_056 (based on primer set crAss056) and CPQ_064 (based on primer set crAss064) qPCR assays was evaluated through a series of paired experiments with established HF183/BacR287 and HumM2 assays. The crAssphage-based assays exhibited high calibration model performance, comparable to the performance of HF183/BacR287 and HumM2 (Table B5). In addition, the crAssphage qPCR genetic markers were present at similar concentrations to HF183/BacR287 and were significantly more abundant ($p > 0.05$) compared to HumM2 in primary influent sewage and impacted environmental water samples tested in this study. In contrast, a recent multiple laboratory evaluation of fecal source identification technologies found sensitivities ranging from 0-60.5% for human-associated viral and bacteriophage genetic markers in challenge samples, in contrast to much higher levels reported for bacterial HF183SYBR (all labs reporting >87% sensitivity) [55, 94]. The high sensitivity exhibited by the crAssphage qPCR assays in this study (100%) is only matched by the pepper mild mottle virus assay [24] and could be another useful alternative to other currently available human-associated viral methods.

In addition to exceptional sensitivity, the crAssphage qPCR assays designed in this study exhibited high specificity (98.6%) based on a fecal reference collection consisting of 222

individual samples from 10 different animal groups. All qPCR assays evaluated in this study exhibited high specificities ranging from 98.6-100%, well above the recommended 80% threshold for water quality management applications [55]. CrAssphage qPCR assays cross-reacted with gull (n=2; 8%) and dog (n=1; 2.4%) samples, both common sources in recreational and residential areas. However, the crAssphage marker concentration was often lower in non-human sources compared to primary influent sewage. In addition, false positives in dog and gull sources was rare, occurring in only 2.4% (1 of 41) dog samples and 8% (2 of 25) gull samples. Other human-associated methods cross-react with these same animal sources likely due to cohabitation with dogs and animal food scavenging [23, 24, 61]. HF183/BacR287 did not cross-react with any samples in the quantifiable range; however, it has been shown to cross react with chicken and turkey at a much lower concentration than in sewage [17]. Despite the high specificity performance of the crAssphage qPCR assays, it is recommended that specificity is confirmed with reference samples from the local area of interest before implementation.

4.4.3 Fecal Source Identification and the Human Fecal Viral Metagenome

This study demonstrates that viral metagenomes are a valuable source of genetic information that can be mined for host-associated sequences to develop novel fecal source identification technologies; however, using metagenomic sequences for method development presents several challenges. First, compiled genomes are constructed from viral quasispecies, resulting in a genome sequence with unknown variability and stability. In addition, novel genomes discovered through metagenomic sequences may lack homology with known annotated genes, resulting in poor quality sequence annotation within compiled genomes. Hence, it is difficult to infer specificity of these sequences *in silico* without laboratory testing. To overcome these challenges,

we performed laboratory testing of select regions of the crAssphage genome to find the most abundant and broadly distributed human-associated genetic regions. This approach builds off of bioinformatic predictions, with extensive end-point PCR laboratory screening to narrow down regions for future genetic marker development. This strategy could also be used for other human-associated viruses; for example, bacteriophages that infect *Bacteroides* strain GB-124 [125, 126]. Currently, no qPCR assays are available for these phages and they may require the isolation of new hosts based on geographic distribution [127]. This approach will continue to be of use as viral metagenome mining continues to improve with additional research efforts leading to more publicly available datasets.

4.4.4 CrAssphage Fecal Source Identification Application

Findings in this study highlight the benefits of crAssphage qPCR methods for human fecal source identification. First, the abundance of the crAssphage markers in sewage and polluted environmental waters implies it will be possible to monitor in smaller sample volumes ($\leq$ 200mL) compared to typical virus assays requiring $\geq$1 liter. The isolation strategy used in this study allows for simultaneous recovery of bacterial and viral genetic markers, as well as the same DNA purification technique because the crAssphage genome is dsDNA. In addition, there is some evidence of genetic stability for the crAssphage method genome regions based on Sanger sequencing in this study, further showing the potential utility of these assays. Findings also indicate that the crAssphage qPCR assays possess a strong human host association (>98%), performing on par with top bacterial human fecal source identification methods. Lastly, as a viral genetic marker, the crAssphage qPCR assays could be a convenient tool to compliment bacterial fecal pollution monitoring tools in future studies.

Despite the high performance observed with the crAssphage qPCR assays, necessary developments remain prior to application of these methods. In this study, environmental samples were processed with ultrafiltration, which is expensive and time consuming. The concentrating pipette procedure used worked well but was only tested with a small number of samples. More research should be conducted to determine the best concentration strategies for crAssphage, including how matrix composition influences recovery efficiencies. In addition, the crAssphage qPCR assays were found to be 98.6% human-associated. This requires specificity testing to be completed in each geographic region prior to implementation, especially in areas with high densities of dogs or gulls. Due to this cross-reaction, it may be necessary to pair these methods with other established human fecal identification technologies in a toolbox approach to improve confidence in results. In addition, future studies may be needed to verify the temporal genetic stability of the crAssphage DNA target sequences, even though the results of this study suggest some level of conservation. Additional studies are necessary to understand linkages of the crAssphage methods to currently recommended fecal indicators, other fecal source identification targets, and pathogens with public health relevance. Lastly, the bacterial host and genome sequence variability of crAssphage remains unconfirmed. While this information is not required to exploit crAssphage for fecal source identification, this information could prove valuable to further utilize this virus for other water quality management applications. The availability of reliable viral assays that are abundant in impacted waters could have broad implications for water quality monitoring and human fecal waste treatment.

5.0 EVALUATION OF CRASSPHAGE-BASED QPCR MARKERS IN AN IMPACTED URBAN WATERSHED

Environmental waters are monitored for fecal pollution to protect the public during recreational activities as well as protecting waters used for drinking sources or food production. While many human-specific source tracking markers have been developed, many lack adequate sensitivity to be reliably detected in environmental waters or do not correlate well with viral pathogens. Recently, two novel human-associated source tracking qPCR markers based on the bacteriophage crAssphage, CPQ_056 and CPQ_064, were developed. These assays were highly human specific, abundant in sewage, and are viral-based, improving on many existing technologies. A 30-day sampling study was conducted in an urban stream impacted by combined sewer overflows to evaluate the crAssphage markers performance in an environmental system. The crAssphage assays were present at concentrations of 4.02-6.04 $\log_{10}$ copies/100mL throughout the study period, indicating their high abundance and ease of detection in polluted environmental waters. In addition, the crAssphage assays were correlated with rain events, molecular markers for human polyomavirus and HF183, as well as culturable *E. coli*, enterococci, and somatic coliphage. Stronger correlations were demonstrated for the crAss_064 assay compared to the crAss_056 assay. This study is the first to demonstrate extended environmental application of crAssphage markers for monitoring of environmental waters.

5.1 INTRODUCTION

The release of inadequately or untreated wastewater into the environment is a major source of microbial impairment of environmental waters. Release of viral pathogens occurs from WWTP effluent, where viruses are not eliminated as successfully as bacteria through treatment processes and can often pass through the system.[86, 87, 172] Another source of pollution comes from leaky septic tanks which have been shown to be the main source of fecal bacteria into the environment in some watersheds.[173] In addition, approximately 40 million people in the United States in 772 communities are serviced by combined sewer systems [174], which are engineered to overflow untreated sewage into natural water bodies during wet weather events, allowing viruses and bacteria from fecal material to pose a health risk to the public.

Fecal contamination in environmental waters is regulated through monitoring of fecal indicator bacteria (FIB). Well-known limitations of monitoring environmental waters with FIB include the lack of differentiation between sources of fecal contamination and inadequately capturing viral risk to human health. The field of microbial source tracking (MST) has emerged to address the former problem, aiming to develop host-specific assays for water monitoring applications. Assays have been developed for the detection of numerous animal species such as cattle [12], dog [11], and gull [10], as well as human-derived fecal matter.[15, 17, 21, 23, 24] Human-specific markers are particularly desirable since this pollution has been shown to pose greater health risks to humans than pollution from other animal sources.[18, 19, 52]

In addition to the inability to differentiate animal sources of fecal pollution, FIB have been shown to not correlate well with enteric viruses in environmental waters and through wastewater treatment processes.[1, 87, 172] Enteric viruses cause a significant amount of disease burden and cost society billions of dollars every year in health system costs. For example, total

norovirus infection costs society an estimated $4.2 billion in direct health systems costs and $60.3 billion in societal costs annually worldwide.[3] In addition to person-to-person contact, the public can be exposed to enteric viruses present in environmental waters through direct contact with impaired waters or through irrigating crops with contaminated water. It has been shown that approximately 16% of all norovirus gastroenteritis illness is caused by environmental sources, such as consuming contaminated water and shellfish.[7] In addition, the occurrence of outbreaks caused by enteric viruses is rising.[1, 3, 4] Though environmental waters are managed in order to reduce the public's exposure to health risks, no human-specific viral-based marker is currently widely used due to low or variable concentrations in sewage, resulting in low detection rates and necessitating concentration of large volumes of water.[94, 107] Viral concentration methods often exhibit low recovery efficiencies and can co-concentrate inhibitors that interfere with downstream testing, such as PCR.[26-28] Many researchers have emphasized the need for a human-associated viral marker that is abundant and reliably detected.[8, 87, 155]

The bacteriophage crAssphage was discovered by metagenomic data mining and found to be highly human-associated and more abundant than other human gut phages and viruses.[32] CrAssphage was suggested for MST technology development and the crAssphage-based qPCR assays used in this manuscript, named CPQ_056 and CPQ_064, were recently published.[164, 175] Other primers not evaluated in this manuscript have also been recently published.[176, 177] Results in the initial publication indicated that the CPQ_056 and CPQ_064 targets were as abundant as current bacterial-based assays in sewage and had high specificity to sewage.[175] The initial study also confirmed crAssphage presence in an environmentally-impacted sample.[175] A recent study evaluating CPQ_056 also confirmed high abundance in sewage and successful detection in a storm sewer outfall.[178] Despite prior successful short-term

demonstrations, environmental assay testing over an extended study period is necessary to evaluate assay performance and determine the environmental correlation of crAssphage markers with existing markers and pollution events.

The goals of the present study were to perform an environmental evaluation of crAssphage qPCR markers in an impacted urban stream and compare assay performance to other culturable and molecular markers. Nine Mile Run in Pittsburgh, PA was monitored daily for 30 days for chemical and biological indicators of water quality. Ranges of concentrations for MST markers were measured and Spearman's rank correlation was determined between all monitored parameters and assays to evaluate crAssphage performance. This is the first extended environmental validation study of the usefulness of crAssphage-based assays for source tracking of human fecal pollution, correlating the assays with culturable indicators, viral indicators, molecular indicators, and pollution events.

5.2 MATERIALS AND METHODS

5.2.1 Study Site

Sampling for the present study was conducted at Nine Mile Run in Pittsburgh, PA daily from September 6, 2016 through October 5, 2016. The selected location has been previously used as a water quality study site.[179] Nine Mile Run is a small urban stream located in Pittsburgh's Frick Park, with a watershed of approximately 19.4 square kilometers.[180] Nine Mile Run has been shown to be chronically polluted, including high levels of fecal indicator bacteria (*E. coli* and fecal coliforms), despite undergoing a restoration project in 2006.[180, 181] The small

stream can become overwhelmed during large rain events, and two combined sewer outfalls discharge directly into the stream. The specific sampling location (40.426385, -79.905262) used in this study is directly downstream of one combined sewer outfall in order to capture changes in concentration to the stream from this point source (Figure 5.1).

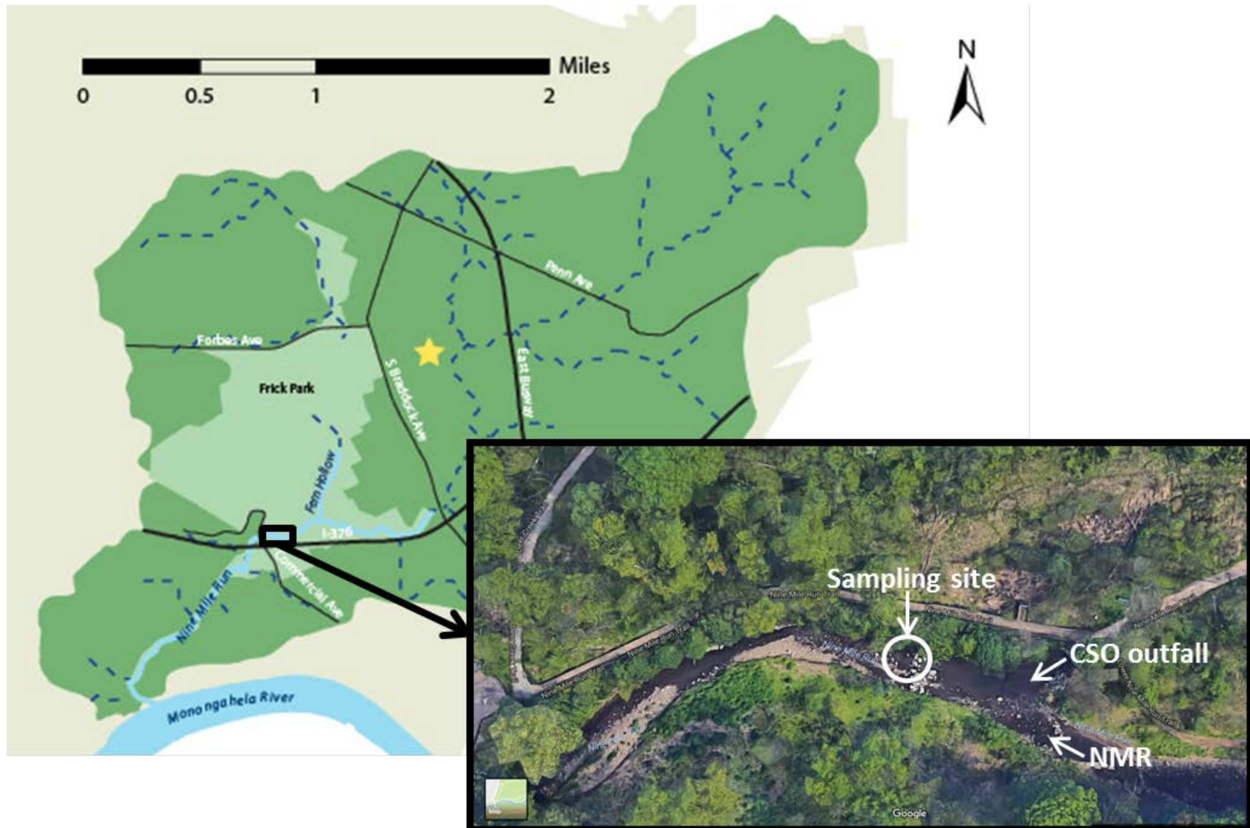

Figure 5.1: Depiction of study sampling site within Frick Park. Google Map image (Map data © 2017 Google) shows the sampling location downstream of where a CSO outfall joins with Nine Mile Run (NMR).

5.2.2 Sampling and Data Collection

Samples were collected daily at 9:00 AM for 30 days. Each day, three liters of stream water were collected in sterile containers and transported on ice to the laboratory, where samples were processed within six hours. In addition, one additional sample was collected on the September 30th during an active CSO event. The temperature of both the stream and the air were measured at the time of sampling. Rainfall data was collected in real-time from 3 Rivers Wet Weather (http://www.3riverswetweather.org/municipalities/calibrated-radar-rainfall-data), a nonprofit organization that operates and maintains rain gauges throughout the region. The data used for this project comes from rain gauge #11, which corresponds to Nine Mile Run. Combined sewer overflow (CSO) events were monitored from the local wastewater treatment and conveyance authority's (Alcosan) online Sewer Overflow Advisory Key status changes (http://www.alcosan.org/SewerOverflowAdvisories/SOAKStatusChanges2017/tabid/115/Default.aspx). A CSO advisory is issued based on the wet well elevation at Alcosan's Main Pump Station, which may or may not correspond to a CSO at Nine Mile Run.

5.2.3 Chemical Parameter Characterization

Chemical parameters were measured throughout the course of this study to monitor water quality. Sample pH was measured using a FiveEasy Plus pH meter (Mettler Toledo, Columbus, OH). Turbidity was measured using Program 745 on a Hach DR900 meter (Loveland, Colorado). Total Organic Carbon (TOC) was measured by filtering 40 mL of sample through a 0.45µm filter and analyzing on a TOC-L TOC Analyzer (Shimadzu, Kyoto, Japan). Total dissolved solids (TDS) were determined by filtering the sample through a 0.45µm filter and

drying in a lab oven at 180°C according to EPA Method 8163. All chemical parameters were tested in triplicate for each sample.

5.2.4 Enumeration of Culturable Indicators

Culturable bacteria and phage indicators were measured for each sample. *E. coli* were measured by filtration and culturing on HiChrome m-TEC agar (Sigma-Aldrich, St. Louis, MO) following the protocol of EPA Method 1603 [182]. Enterococci were measured by filtration following EPA Method 1600 and culturing on mEI plates [183], made with Difco mE agar (BD, Franklin Lakes, NJ) with the addition of indoxyl β-D-glucoside (Chem-Impex International, Wood Dale, IL). Somatic coliphages were enumerated via a single agar layer plaque assay procedure according to EPA Method 1602 [163].

5.2.5 qPCR Assays

Stream water was simultaneously concentrated for viruses and bacteria as previously described.[23, 28] Briefly, 500 mL of stream water was pH adjusted to pH=3.5 and filtered through a 47mm 0.45µm mixed cellulose HAWG filter (Millipore, Billerica, MA). Duplicate filters for each sample were then placed in tubes and frozen at -80°C for later bulk DNA extraction. For DNA extraction, filters and bead tubes were allowed to thaw to room temperature and DNA was extracted using a DNeasy PowerWater Kit (Qiagen, Valencia, CA), following manufacturer's protocol. Extracted DNA was stored at -20°C. TaqMan qPCR assays targeting crAssphage (CPQ_056 and CPQ_064),[175] a *Bacteroides* species (HF183/BacR287),[17] and human polyomavirus[23] (HPyV) were performed as described previously. Each 25 µL qPCR

reaction for the crAssphage assays and HF183/BacR287 contained 1x TaqMan Environmental Master Mix 2.0 (Thermo Fisher Scientific), 5 µg Bovine Serum Albumin (BSA, Sigma-Aldrich, St. Louis, MO), 1 µM of each primer, 80 nM FAM-labeled probe, 80 nM VIC-labeled probe (HF183/BacR287 only), molecular grade water, and 2 µL of a 10x dilution of extracted DNA. Each 25 µL reaction for HPyV contained 1x TaqMan Environmental Master Mix 2.0, 5 µg Bovine Serum Albumin, 0.5 µM of each primer, 0.4 µM FAM-labeled probe, molecular grade water, and 2 µL of extracted DNA. All qPCR tests were performed in triplicate on a CFX Connect Real-Time System (BioRad, Hercules, CA). Each qPCR plate included a standard curve with 10^1-10^5 target copies per reaction that was prepared as previously described.[175] Thermocycler conditions for each assay were as previously reported.[23, 175] Fluorescence drift correction was applied and the threshold for quantification cycle (C_q) determination was auto calculated using a baseline determined from cycle numbers 10 to 20.

5.2.6 Controls

Three filter control DNA extractions were performed by placing a sterile filter in a bead mill tube, and three extraction blanks were performed and tested by qPCR for all assays. Controls for mTEC and mEI plates were performed by filtering sterile buffer rinse water through a filter and plating. Controls for the somatic coliphage procedure were performed by adding DI water to the test agar. Triplicate no template controls (NTCs) were included on each qPCR plate. All no template controls and process controls were negative throughout the course of the study. In addition, the HF183/BacR287 assay includes an internal amplification control (IAC) that allows for evaluation of PCR inhibition. Potential inhibition introduced through the water environment

or concentration procedures was determined as previously described.[175] No samples exhibited PCR inhibition as determined through the evaluation of the IAC control.

5.2.7 Statistical Analysis

Spearman's rank correlation coefficients (r) were calculated on the means of log transformed data in GraphPad Prism 7 (La Jolla, CA) between culturable indicators, qPCR indicators, and chemical parameters, using two-tailed 95% confidence intervals. Coefficients are characterized by the following scale for comparison purposes: 0.2-0.39 (weak correlation), 0.4-0.59 (moderate correlation), 0.6-0.79 (strong correlation), and 0.8-1 (very strong correlation). One-way ANOVAs were performed in Minitab 17 (State College, PA) on $\log_{10}$ transformed data between dry weather days and wet weather days to determine if rain events significantly impacted water quality parameters. Rainfall is reported as the total amount of precipitation recorded in the 24 hours before sampling. A sampling day was considered to be a wet weather sampling event if it had rained within the past 24 hours.

5.3 RESULTS AND DISCUSSION

5.3.1 Site Background

The goal of the current study was to demonstrate the use of crAssphage based qPCR assays to detect and quantify human fecal pollution in an impacted urban watershed. The study was conducted at Nine Mile Run in Pittsburgh, PA. Nine Mile Run is a small urban waterway

perpetually impacted by fecal contamination through combined sewer overflow events and storm water runoff. This test site was chosen for this study due to known contamination events from CSO events and leaky sanitary sewers[179], facilitating detection and evaluation of fecal source tracking markers in a relevant environmental matrix. Samples were collected from a public access location point daily for a 30 day period in September 2016. An additional sample was also collected during an active CSO event. Samples were then transported to the laboratory at the University of Pittsburgh and immediately processed for water quality parameters and culturable indicators.

5.3.2 Rainfall During the Study Period

During the 30-day study period, it rained a total of 12 days, with 8 of those days reporting active CSOs. Total rainfall ranged from 0.02 to 1.01 inches of rain in a 24-hour period.

5.3.3 Chemical Parameters During the Study Period

Chemical parameters were measured throughout the length of the study to monitor water quality. Results for TOC, pH, turbidity, and TDS can be seen in Figure 5.2 along with a plot of the rainfall throughout the study. Throughout the study period, TOC ranged from 1.6-6.3 mg/L (1.6-3.4 mg/L for dry weather days and 2.4-6.3 mg/L for wet weather days). The pH remained stable throughout sampling, ranging from 7.62 to 8.05. Turbidity ranged from non-detect (ND) to 40.7 Formazin Attenuation Units (FAU) (ND-10 FAU for dry weather days and 2.7-40.7 FAU for wet weather days). TDS ranged from ND-947 mg/L (500-947 mg/L for dry weather days and ND-692 mg/L for wet weather days). The chemical values measured during the active CSO event fell

within the range observed on study days for TOC and TDS but were observed outside of the wet weather range for pH (6.73) and turbidity (87.3 FAU).

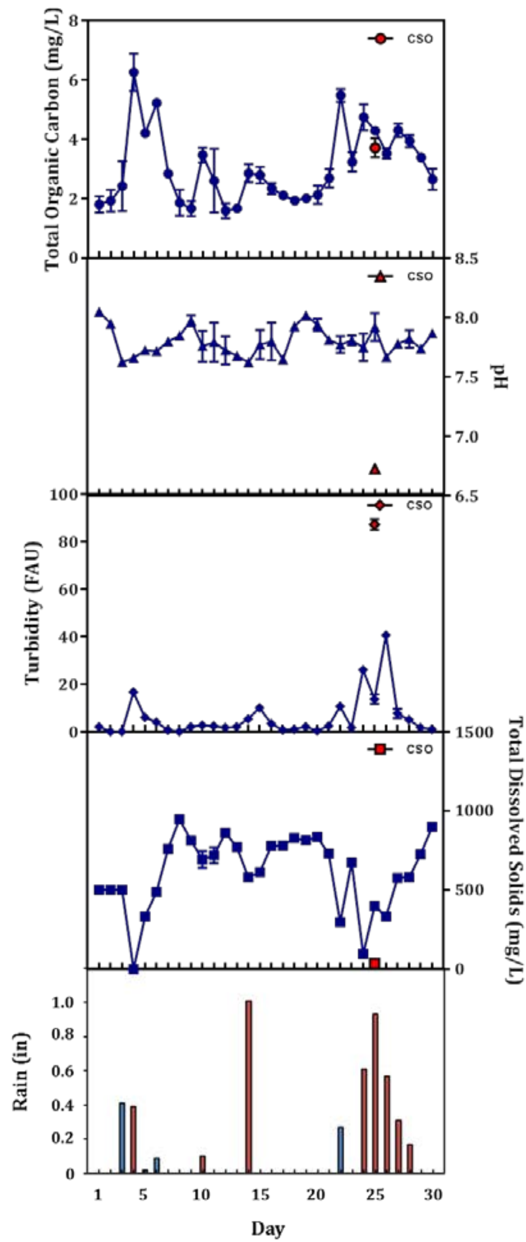

Figure 5.2: Chemical parameters measured along with daily rainfall during the Nine Mile Run sampling period of 30 days. Rainfall is presented as the amount of rain that fell within the 24 hour period before that day's sampling. A red bar on the rainfall graph indicates rainfall that corresponded to a reported CSO event. The individual data points shown in red represent an additional sampling time point during day 25 during an active CSO event.

5.3.4 Correlation of Chemical Parameters with Rainfall Events

One-way ANOVA analysis was performed between observed chemical values on dry weather days versus wet weather days. The analysis demonstrated a statistically significant difference ($p < 0.05$) for all chemical parameters between dry and wet weather days. TOC and turbidity were both positively correlated with rain events, i.e. their values increased significantly with rainfall. TDS and pH were both negatively correlated with rainfall, i.e. these parameters were characterized by significantly lower values on wet weather sampling days. No significant differences in values were observed between wet weather events that resulted in a reported CSO event versus no reported CSO event.

5.3.5 Culturable and qPCR Indicators During Study Period

Culturable and qPCR assays were evaluated in order to compare the crAssphage qPCR assays to fecal pollution indicators (Figure 5.3). Ranges observed for culturable indicators were: *E. coli* (2.38-4.49 $\log_{10}$ cfu/100mL), enterococci (2.06-4.31 $\log_{10}$ cfu/100mL), and somatic coliphage (2.02-3.74 $\log_{10}$ pfu/100mL). Ranges observed for qPCR indicators were: CPQ_056 (4.02-6.04 $\log_{10}$ copies/100mL), CPQ_064 (4.33-5.94 $\log_{10}$ copies/100mL), HF183/BacR287 (2.80-5.64 $\log_{10}$ copies/100mL), and HPyV (1.50-4.21 $\log_{10}$ copies/100mL). All indicators were detected each day of the study except for human polyomavirus (HPyV) on a single sampling day.

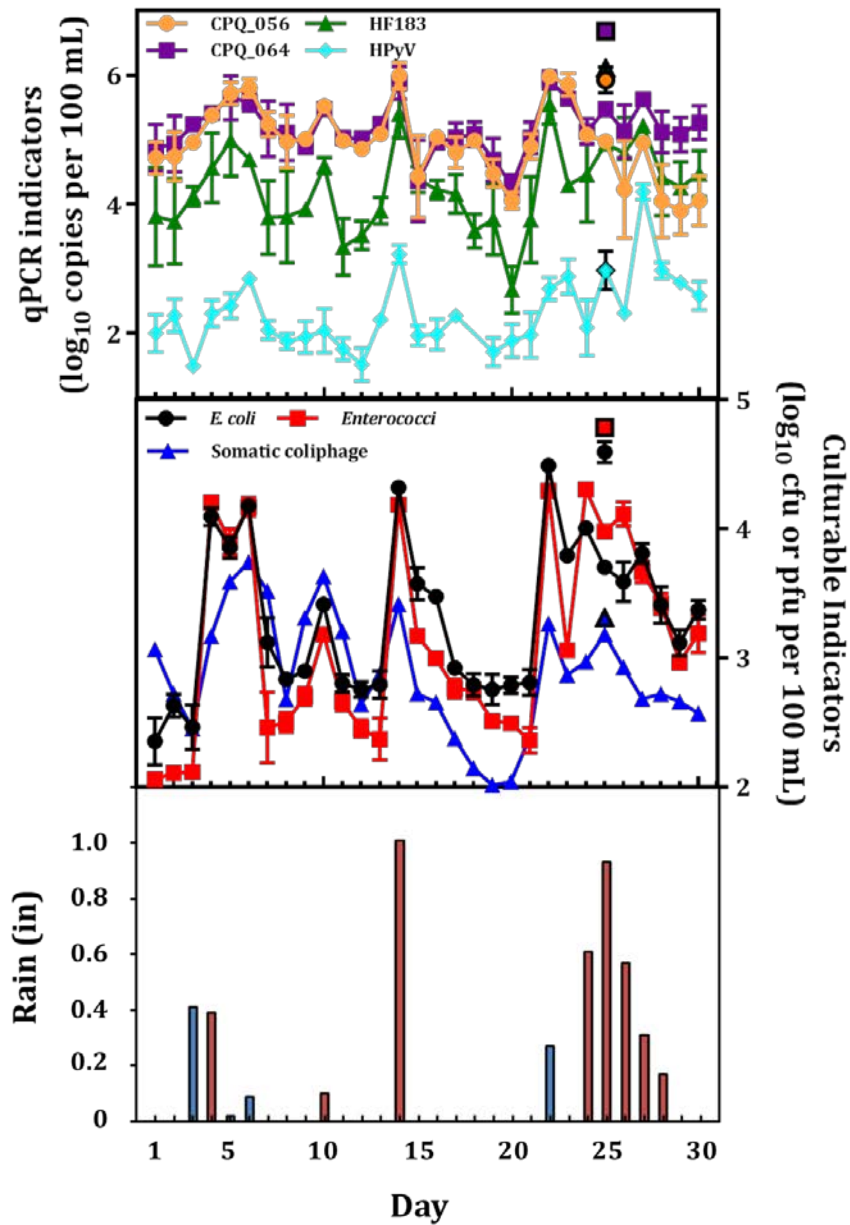

Figure 5.3: Culturable and qPCR indicators along with daily rainfall during the Nine Mile Run sampling period of 30 days. Rainfall is presented as the amount of rain that fell within the 24 hour period before that day's sampling. A red bar on the rainfall graph indicates rainfall that corresponded to a reported CSO event. The individual data points for the indicators represent an additional sampling time point during day 25 during an active CSO event.

5.3.6 Correlation of Human-associated Markers with Rainfall Events

Similarly to the chemical parameters, one-way ANOVA analysis was performed on all biological indicators to understand if rain and/or CSO events significantly affected concentrations of the markers. The abundance of all markers was positively correlated with rain events, i.e. in all cases the indicator increased after a rainfall event. In addition, the abundance of all markers was found to be statistically significantly different ($p < 0.05$) between wet and dry weather conditions. No statistically significant difference was found between wet weather events that corresponded to a reported CSO event versus a rain event without a reported CSO event.

5.3.7 Correlation of chemical and biological water quality parameters

Spearman's rank correlation coefficients were calculated between each pair of culturable indicators, molecular indicators, and chemical parameters (Figure 5.4, Table C1). The reported correlation coefficients generally showed strong correlation between markers, with 91% of p-values being statistically significant (<0.05) (Table C2). Although the two crAssphage assays strongly correlated with each other, their correlation with other indicators varied. The only chemical parameter that correlated strongly with crAssphage abundance was TOC with the CPQ_064 assay. pH did not correlate strongly with any other biological or chemical parameter. For culturable markers, the CPQ_064 assay strongly correlated with culturable *E. coli* and enterococci, while the CPQ_056 assay correlated strongly with somatic coliphage. For molecular markers, the CPQ_056 assay was only strongly correlated with CPQ_064. In contrast, the CPQ_064 assay was strongly correlated with the HF183 and HPyV assays, in addition to CPQ_056.

In addition to the crAssphage correlations described above, there were several other correlations between other parameters measured in this study. The strongest correlations were observed between bacterial based indicators: *E. coli* vs. enterococci (0.93), *E. coli* vs. HF183 (0.88), and enterococci vs. HF183 (0.86). The bacterial based indicators (*E. coli*, enterococci, and HF183) also had stronger correlations with chemical parameters than the viral indicators exhibited, with each being strongly correlated with TOC, turbidity, and TDS.

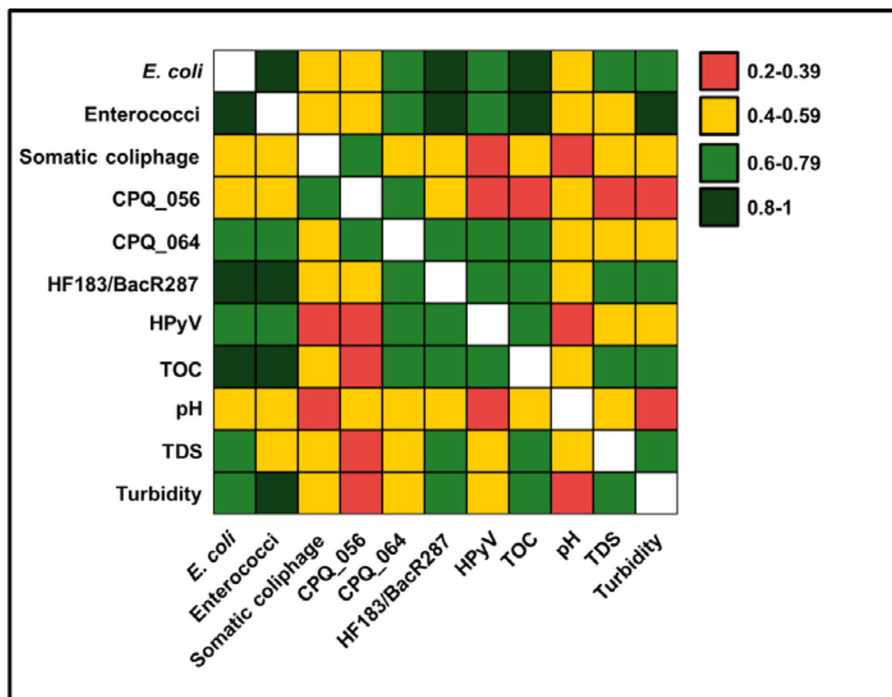

Figure 5.4: Heat map of Spearman's rank correlation coefficients matrix for culturable indicators, qPCR indicators, and chemical parameters. Coefficients are colored based on the following scale of the absolute value of the coefficients: red (0.2 to 0.39 - weak correlation), orange (0.4 to 0.59 - moderate correlation), green (0.6 to 0.79 - strong correlation), and dark green (0.8 to 1 - very strong correlation). Actual values of correlation coefficients are provided in Table C1.

5.3.8 Comparison of crAssphage qPCR Assays

Overall, levels of both crAssphage markers were detected at similar concentrations to HF183/BacR287 and approximately two to three orders of magnitude greater than HPyV. Despite positive correlations between all biological indicators, the two crAssphage-based assays differed in the strength of their correlations with other indicators. The most notable difference is the strong correlation between the CPQ_064 assay and the HPyV assay, while the p-value between CPQ_056 and HPyV was one of the few in the study to not show statistical significance. Differences in assay performance highlighted in this study are notable and warrant further investigation, as it is likely that future investigations will choose a single crAssphage assay to apply, for example a recent study investigated only CPQ_056.[178]

This study demonstrates the high abundance of crAssphage-based indicators compared to other indicators of human fecal pollution in the environment and the ease at which they can be detected along with other DNA-based assays. Due to the high levels of fecal pollution at the selected study site, HPyV was detected in nearly all samples; however, HPyV had the lowest quantities detected of all the qPCR assays. Based on the results of this study, CPQ_064 correlates well with HPyV yet is orders of magnitude more abundant, making crAssphage assays potentially useful in waters polluted at lower levels.

Recently, a separate crAssphage-based qPCR assay not evaluated in the current study was developed by targeting a short (78bp) genetically conserved genome fragment and evaluated.[176] In the initial evaluation of this assay, 61% of fecal samples derived from cows, pigs, and poultry were positive for this crAssphage sequence.[176] In comparison, the assays evaluated in this study have been shown to exhibit specificities of 98.6% (n=222), cross-reacting with a single dog fecal sample and two gull samples.[175] A recent study also showed CPQ_056

cross-reaction with poultry litter but still determined a specificity of 92.7%.[178] This further demonstrates the differing suitability of crAssphage genome regions for human-specific applications. Some genome regions of crAssphage may have homology to other phages within other animals' intestinal tracts, as well as some regions may be human-specific. Hence, it is likely that crAssphage assays will perform differently based on the genomic region on which they are designed.

5.3.9 Site Specific Considerations

While the goal of this study was to demonstrate performance of crAssphage qPCR assays in an impacted urban stream whose results can be applied widely, the specific site chosen has some unique characteristics that should be considered. Nine Mile Run was selected as a test site due to its well-known contamination due to sewage inputs and the well-characterized nature of the system. The present study detected peaks in biological marker concentrations after rain events, which then decayed to a background level of pollution. Ideally, the peak in concentration would decay until the biological markers are no longer detectable to confirm that the marker signal is not present if fecal pollution is no longer present. However, NMR has been shown to be impacted by leaking sanitary sewers during baseflow conditions,[179, 184] resulting in a signal decay to this lower level of constant sewage input. Additional research should be conducted to quantify the decay of the crAssphage markers. In addition, it has been shown that during rain events, more pristine storm water dilutes the nitrate signal in the stream.[179, 184] This effect could also be present in this study, muting the peak concentrations of the biological markers.

The current study evaluated a single system and results may vary between environmental systems. Previous studies have suggested the potential for geographic variation of crAssphage

abundance in sewage;[164] the crAssphage indicators should be studied in additional sites to verify findings. In addition, the study site demonstrated high levels of microbial water quality impairment. The concentrations of *E. coli* and enterococci were above the 2012 Recreational Water Quality Criteria recommendations throughout the duration of our study;[185] however, elevated levels of these indicator bacteria were readily observed on wet weather days allowing observations above background levels of pollution. Future studies should investigate both systems less chronically polluted and systems polluted by non-point sources to evaluate if crAssphage-based markers correlate better than indicator bacteria to pollution events in systems that are impacted from additional non-human pollution sources. Further, this study found no significant differences from CSO events and rain events not resulting in a reported CSO. This could be due to limitations in data resolution since the study system is located away from where the CSO levels are determined. Overflow events may have been occurring even when they were not officially reported. In addition, overflow events may have overwhelmed the study stream faster than the reporting system due to the small volume of the study stream. Finally, samples were necessarily concentrated for qPCR analysis. The concentration method used in this study conveniently allowed for simultaneous concentration of both viral and bacterial indicators. This method has previously been shown to perform well, with HPyV recovery efficiencies of 40 and 78% in tap and river water, respectively;[28] however, the use of any concentration method likely underestimates the levels of both crAssphage and human polyomavirus.

5.3.10 Environmental Implications

This study demonstrated the first extended environmental application of crAssphage-based qPCR assays as indicators of human fecal pollution in an impacted urban stream. The crAssphage

assays were shown to correlate with other culturable and molecular indicators, as well as rain events. The assays were also shown to be present in contaminated waters at high abundance, orders of magnitude higher than the viral-based HPyV assay. This suggests their ease of detection and utility in less chronically impaired waters where other viruses may be too dilute to detect. While crAssphage-based assays remain a promising tool for water management applications, more research is necessary to fully understand marker performance. Additional study sites should be evaluated that are affected by non-point sources of pollution and that are less chronically polluted than Nine Mile Run. In addition, studies should be conducted to understand the persistence of crAssphage marker signals in different water matrices compared to other pathogenic viruses to understand if crAssphage behaves in a similar manner. Ultimately, these developments will improve our ability to more accurately assess viral presence and health risk in waters to ensure a more complete picture of water quality composition.

6.0 SUMMARY AND CONCLUSIONS

6.1 MAJOR FINDINGS

Environmental water quality monitoring is imperative to protect both environmental health and public health. A review of current microbial source tracking technologies showed that current techniques do not adequately capture viral health risk and are unable to distinguish the animal source of the pollution (Chapter 2). MST efforts have produced many new technologies for biological screening of environmental waters based on bacteria, viruses, and bacteriophages. Challenges remain with all assay types. Bacteria based MST assays cross-react with non-human animal fecal matter and do not correlate with viral presence in environmental waters. Viral assays are less abundant in environmental waters than bacteria and phages, making their detection unreliable and necessitating concentration methods that can be inefficient or increase inhibition in downstream testing. Bacteriophage assays largely remain culture based, which can necessitate additional bacterial host isolation and can cross-react with other animal species similarly to their bacterial hosts. The shortcomings of existing technologies were highlighted and ideal assay characteristics were identified.

The bacteriophage "crAssphage" was identified as a candidate for a human-specific source tracking technology based on the criteria established in Chapter 2. A metagenomic evaluation of crAssphage for a source tracking specific application was performed (Chapter 3).

CrAssphage was abundant in human-derived sewage, although geographic variability in crAssphage abundance was observed. While crAssphage was observed in human sewage, it was not observed in other environments. This implies crAssphage is largely associated with the human gut. In addition, crAssphage was found to be more abundant than other viruses that have been proposed as viral water quality indicators. This metagenomic screening provided an economic and time efficient method of initial screening of a novel MST technology. The high abundance and human association demonstrated by crAssphage in this study warranted further investigation of crAssphage as an MST target.

In Chapter 4, primer pairs were designed along the length of the crAssphage genome, tested by three rounds of end-point PCR, and then the best-performing primer pairs were adapted to qPCR assays. A total of 48% of the crAssphage genome was initially identified as ideal for primer development, eliminating regions suspected of being hyper-variable or non-selectively hitting to other species targets. Primer design resulted in 384 primer pairs spanning the crAssphage genome that were subjected to three rounds of PCR testing. After rigorous testing by PCR, two primer pairs were adapted to qPCR assays, CPQ_056 and CPQ_064. These assays were tested head-to-head with two bacterial-based assays and were found to be just as abundant or an order of magnitude more abundant than the bacterial-based assays. In addition, the assays were found to exhibit high human-specificity. This research detailed the development of novel viral-based qPCR technologies that are both highly human-associated and highly abundant in reference materials, improving on the lack of sensitivity experienced with existing viral technologies.

An initial environmental evaluation of the developed crAssphage assays was performed (Chapter 5). Results showed that the crAssphage assays were significantly correlated with rain

events. In addition, the assays were positively correlated with culturable indicator bacteria and bacteriophages as well as with molecular bacterial and viral markers. However, this study demonstrated varied assay performance, suggesting the two developed crAssphage assays may perform differently based on the genomic area on which they are designed. This study further demonstrated the high abundance of crAssphage in environmental waters and provided evidence for assay correlation with existing technologies and pollution events.

6.2 IMPLICATIONS

The novel technologies developed from this research effort contribute to the tools available for monitoring environmental waters. Currently, environmental waters are monitored using general bacterial indicators of fecal pollution. These general indicators do not differentiate pollution based on its source and do not necessarily correlate with viral health risk in polluted waters. Two qPCR assays, CPQ_056 and CPQ_064, were developed herein. The developed assays improve upon existing technologies by: (1) exhibiting high human-specificity, (2) having high concentrations in reference materials, and (3) being of viral origin. CrAssphage-based assays have the potential to better represent viral risk to human health in environmental waters while remaining more reliable due to their ease of detection.

6.3 FUTURE RESEARCH

Despite the contributions to the field from this dissertation research, future research is warranted in further evaluating crAssphage-based technologies for implementation in water management applications. The two assays designed should be verified with additional human and animal fecal reference samples to establish their specificity values over a wider geographic range, in addition to international surveillance. Further environmental surveillance should be conducted, including studies in less impacted sites and from non-point sources of pollution, such as agricultural runoff. Also, multi-lab studies should be conducted to ensure the quality of the crAssphage assays' performance metrics, where they can be tested head-to-head with other MST technologies. Further research effort is required to correlate additional properties of the assays to those of human viral pathogens, including decay rates of crAssphage marker signal in environmental waters as well as through wastewater treatment. Lastly, the bacterial host of crAssphage remains unknown; therefore, there is no method available for culturing and isolating crAssphage in lab. While the research conducted in this dissertation provides an example of new MST technology develop based on bioinformatic data rather than culture-based, culturing methods would be desirable to study basic crAssphage biology, such as persistence in the environment and reaction to disinfectants. Established culturing methods will facilitate knowledge acquisition and implementation of crAssphage-based markers for water quality management.

APPENDIX A

CHAPTER 3 – SUPPLEMENTAL INFORMATION

Table A1: Metagenomes used in this study along with individual sample data and alignment results. References are provided where available.

sample	environment	metagenome type	number of sequences mapped	% sequences mapped	fraction of reference covered	average coverage	sequencing method	accession	reference
raw sewage - Barcelona	human associated	viral	1560	0.15	0.76	5.47	454 GS FLX Titanium	SRA040148	[1]
raw sewage - Pittsburgh	human associated	viral	2178	0.33	0.9	10.06	454 GS FLX Titanium	SRA040148	[1]

Table A1 (continued)

raw sewage - San Francisco	human associated	viral	454	0.08	0.42	1.44	454 GS FLX Titanium	SRA054852	[2]
A_influent - United States	human associated	viral	7948	0.34	0.64	2.27	Illumina HiSeq	provided by author	[3]
A_effluent - United States	human associated	viral	4741	0.19	0.46	1	Illumina HiSeq	provided by author	[3]
B_influent - United States	human associated	viral	8089	0.1	0.78	2.58	Illumina HiSeq	provided by author	[3]
B_effluent - United States	human associated	viral	6721	0.08	0.58	1.5	Illumina HiSeq	provided by author	[3]
C_influent - United States	human associated	viral	4731	0.06	0.61	1.94	Illumina HiSeq	provided by author	[3]
C_effluent - United States	human associated	viral	6508	0.08	0.67	1.75	Illumina HiSeq	provided by author	[3]
D_influent - United States	human associated	viral	12110	0.15	0.82	4.59	Illumina HiSeq	provided by author	[3]
D_effluent - United States	human associated	viral	7585	0.09	0.61	1.79	Illumina HiSeq	provided by author	[3]
E_influent - United States	human associated	viral	10639	0.13	0.72	4.05	Illumina HiSeq	provided by author	[3]
E_effluent - United States	human associated	viral	8411	0.11	0.61	2.04	Illumina HiSeq	provided by author	[3]

Table A1 (continued)

raw sewage - Addis Ababa	human associated	viral	446	0.07	0.06	1.78	454 GS FLX Titanium	SRA040148	[1]
raw sewage - Nigeria	human associated	viral	0	0	-	-	454 GS FLX Titanium	SRA054852	[2]
raw sewage - Nepal	human associated	viral	16	0	0.02	0.04	454 GS FLX Titanium	SRA054852	[2]
raw sewage - Bangkok	human associated	viral	8	0	0.02	0.02	454 GS FLX Titanium	SRA054852	[2]
influent - Singapore	human associated	viral	88	0.05	0.3	0.37	454 GS FLX Titanium	4450219.3	[4]
activated sludge - Singapore	human associated	viral	5	0	-	-	454 GS FLX Titanium	4450221.3	[4]
anaerobic digester - Singapore	human associated	viral	10	0	-	-	454 GS FLX Titanium	4450224.3	[4]
raw sewage - China	human associated	viral	4537	0.01	0.87	4.79	Illumina HiSeq 2500	SRX450092	
fecal matter - California sea lion	terrestrial animals	viral	0	0	-	-	454 GS FLX Titanium	SRA044033	[5]
fecal matter - wild rodent	terrestrial animals	viral	0	0	-	-	454 GS FLX Titanium	SRA030869	[6]

Table A1 (continued)

bovine rumen	terrestrial animals	viral	4	0	-	-	454 GS FLX Titanium	SRX215440	
fecal matter - pig	terrestrial animals	viral	0	0	-	-	454 GS FLX Titanium	SRA030664	[7]
bat guano	terrestrial animals	viral	0	0	-	-	454 GS FLX Titanium	SRA012669	[8]
bat guano	terrestrial animals	viral	25991	0.3	0.39	5.41	Illumina Genome Analyzer	DRA000500	[9]
cow rumens pool plankton	terrestrial animals	microbial	1	0	-	-	454 GS20	4440357.3	analyzed in [16]
cow rumens 80F6	terrestrial animals	microbial	0	0	-	-	454 GS20	4440356.3	analyzed in [16]
cow rumens 640F6	terrestrial animals	microbial	7	0	-	-	454 GS20	4440355.3	analyzed in [16]
cow rumens 710 F	terrestrial animals	microbial	3	0	-	-	454 GS20	4440387.3	analyzed in [16]
lean mice	terrestrial animals	microbial	0	0	-	-	454 GS20	4440463.3	[10], analyzed in [16]
obese mice	terrestrial animals	microbial	0	0	-	-	454 GS20	4440464.3	[10], analyzed in [16]

Table A1 (continued)

chicken cecum NCTC	terrestrial animals	microbial	3	0	-	-	454 GS20	4440367.3	analyzed in [16]
chicken cecum uninfected	terrestrial animals	microbial	5	0	-	-	454 GS20	4440368.3	analyzed in [16]
chicken cecum A	terrestrial animals	microbial	5	0	-	-	454	4440283.3	
chicken cecum B	terrestrial animals	microbial	3	0	-	-	454	4440284.3	
canine beet pulp GI metagenome	terrestrial animals	microbial	0	0	-	-	454	4444702.3	
canine control gastrointestinal metagenome	terrestrial animals	microbial	1	0	-	-	454	4444703.3	
healthy fish slime	fish	viral	0	0	-	-	454 GS20	4440065.3	analyzed in [16]
morbid fish slime	fish	viral	1	0	-	-	454 GS20	4440064.3	analyzed in [16]
mosquito Oceanside Ca	mosquito	viral	0	0	-	-	454 GS20	4440052.3	analyzed in [16]
mosquito San Diego	mosquito	viral	5	0	-	-	454 GS20	4440053.3	analyzed in [16]

Table A1 (continued)

mosquito Mission Valley Ca	mosquito	viral	5	0	-	-	454 GS20	4440054.3	analyzed in [16]
effluent - Singapore	reclaimed water	viral	2	0	-	-	454 GS FLX Titanium	4450223.3	[4]
park - RW	reclaimed water	viral	0	0	-	-	454 GS FLX	SRA008294	[11]
nursery - RW	reclaimed water	viral	0	0	-	-	454 GS FLX	SRA008294	[11]
nursery - RW	reclaimed water	viral	0	0	-	-	454 GS FLX	SRA008294	[11]
effluent - RW	reclaimed water	viral	0	0	-	-	454 GS FLX	SRA008294	[11]
effluent - RW	reclaimed water	viral	1	0	-	-	454 GS FLX	SRA008294	[11]
Solar Salterns medium salinity west California	hypersaline	viral	0	0	-	-	454 GS20	4440427.3	analyzed in [16]
Solar Salterns medium salinity west California	hypersaline	viral	0	0	-	-	454 GS20	4440428.3	analyzed in [16]
Solar Salterns high salinity West California	hypersaline	viral	0	0	-	-	454 GS20	4440421.3	analyzed in [16]
Solar Salterns low salinity San Diego	hypersaline	viral	0	0	-	-	454 GS20	4440436.3	analyzed in [16]

Table A1 (continued)

Solar Salterns low salinity San Diego	hypersaline	viral	3	0	-	-	454 GS20	4440432.3	analyzed in [16]
Solar Salterns medium salinity west California	hypersaline	viral	0	0	-	-	454 GS20	4440431.3	analyzed in [16]
Solar Salterns medium salinity west California	hypersaline	viral	0	0	-	-	454 GS20	4440417.3	analyzed in [16]
Solar Salterns high salinity West California	hypersaline	viral	0	0	-	-	454 GS20	4440145.4	analyzed in [16]
Solar Salterns high salinity West California	hypersaline	viral	0	0	-	-	454 GS20	4440144.4	analyzed in [16]
Solar Salterns low salinity west california	hypersaline	viral	3	0	-	-	454 GS20	4440420.3	analyzed in [16]
Salton Sea	hypersaline	viral	0	0	-	-	454 GS20	4440327.3	analyzed in [16]
Salton Sea	hypersaline	viral	0	0	-	-	454 GS20	4440328.3	analyzed in [16]
Marine GOM	marine	viral	6	0	-	-	454 GS20	4440304.3	[12], analyzed in [16]

Table A1 (continued)

Marine BBC	marine	viral	1	0	-	-	454 GS20	4440305.3	[12], analyzed in [16]
Marine SAR	marine	viral	6	0	-	-	454 GS20	4440322.3	[12], analyzed in [16]
Line Is Kingman	marine	viral	0	0	-	-	454 GS20	4440036.3	[13], analyzed in [16]
Line Is Christmas	marine	viral	1	0	-	-	454 GS20	4440038.3	[13], analyzed in [16]
Line Is Palmyra	marine	viral	8	0	-	-	454 GS20	4440040.3	[13], analyzed in [16]
Line Is Tabuaeran	marine	viral	6	0	-	-	454 GS20	4440280.3	[13], analyzed in [16]
Tampa Bay Mitomycin C induced	marine	viral	6	0	-	-	454 GS20	4440102.3	analyzed in [16]
Skan Bay	marine	viral	1	0	-	-	454 GS20	4440330.3	analyzed in [16]
freshwater lake	freshwater	viral	0	0	-	-	454 GS FLX Titanium	ERP000339	[14]

Table A1 (continued)

freshwater lake	freshwater	viral	0	0	-	-	454 GS FLX Titanium	ERP000339	[14]
tilapia pond 3	freshwater	viral	1	0	-	-	454 GS20	4440424.3	analyzed in [16]
healthy fish pond	freshwater	viral	2	0	-	-	454 GS20	4440412.3	analyzed in [16]
healthy fish prebead	freshwater	viral	0	0	-	-	454 GS20	4440414.3	analyzed in [16]
tilapia pond	freshwater	viral	2	0	-	-	454 GS20	4440439.3	analyzed in [16]
Porites compressa time zero	coral	viral	0	0	-	-	454 GS20	4440376.3	analyzed in [16]
Porites compressa control	coral	viral	0	0	-	-	454 GS20	4440374.3	analyzed in [16]
Porites compressa DOC	coral	viral	1	0	-	-	454 GS20	4440370.3	analyzed in [16]
Porites compressa pH	coral	viral	0	0	-	-	454 GS20	4440371.3	analyzed in [16]

Table A1 (continued)

Porites compressa nutrients	coral	viral	0	0	-	-	454 GS20	4440377.3	analyzed in [16]
Porites compressa Temperature	coral	viral	0	0	-	-	454 GS20	4440375.3	analyzed in [16]
Pozas Azule II	microbialites	viral	1	0	-	-	454 GS20	4440320.3	[15], analyzed in [16]
Rios Mesquites	microbialites	viral	0	0	-	-	454 GS20	4440321.3	[15], analyzed in [16]
Highborne Cay	microbialites	viral	4	0	-	-	454 GS20	4440323.3	[15], analyzed in [16]

Table A2: Number of reads mapped for the U.S. and Europe metagenomes to the viral genomes

	crAssphage	PMMoV	adenovirus	polyomavirus	TTV	norovirus
A_inf	7948	355	365	67	49	65
A_eff	4741	138	374	121	75	103
B_in	8089	583	858	152	118	167
B_eff	6721	317	1007	195	122	169
C_in	4731	62	240	145	36	67
C_eff	6508	229	672	157	96	157
D_in	12110	196	798	182	76	177
D_eff	7585	228	662	197	111	156
E_in	10639	184	1082	138	45	147
E_eff	8411	127	553	226	98	98
San Francisco	454	74	0	0	0	0
Pittsburgh	2178	240	23	0	0	17
Barcelona	1560	2404	2	0	0	0
average	6283	395	510	122	64	102
STD	3456	619	376	80	45	66

APPENDIX B

CHAPTER 4 – SUPPLEMENTAL INFORMATION

Table B1: Summary of reference fecal sample collection

Host (Common Name)	Locality	Count
Cow	OR	52
	WY	9
Dog	OR	11
	WI	11
	OH	10
	FL	9
Gull	OR	5
	WI	10
	OH	10
Horse	OR	11
	WA	9
Elk	OR	11
	WA	9
Canada Goose	OR	10
	OH	8
Chicken	OR	11
Pig	OH	9
Deer	WA	9
Beaver	WA	8
	TOTAL	222

Table B2: Summary of primary influent sewage sample collection

Sample	**Location**
1	Cincinnati, OH
2	Davenport, IA
3	Pittsburgh, PA
4	Keokuk, IA
5	Wichita, KS
6	Miami, FL
7	Corvallis, OR
8	Costa Mesa, CA
9	Milford, OH

Table B3: Candidate primer set sequence information. Information on whether each primer pair passed or failed round I, II, and III testing is given in addition to the reason why each primer pair failed selection criteria.

Assay	Primer	Sequence 5' → 3'	Genomic Region	CrAssphage ORF	Round I	Round II	Round III	Reason for Failure
crAss 001	crAss001-For	GGTGAAATTAGAGGTGACGAGAGG	1772 to 1867	NA	Pass	Fail		Primer Dimer
	crAss001-Rev	CCACCATTTCATTCTCCACGAA						
crAss 002	crAss002-For	GAAGATATTTACGAGCCTTTCCAC	78103 to 78274	77	Fail			Low amplification, primer dimer
	crAss002-Rev	GGTCAGATTACTCCTGAAATGCTT						
crAss 003	crAss003-For	GACTTCCACAAACGTTCTTCGT	83864 to 83966	86	Fail			False Positive
	crAss003-Rev	TCCAAAATGATAATGGTGGTACTTC						
crAss 004	crAss004-For	ACGCATCATAAATAAGAGGATAACG	93164 to 93343	96	Fail			Low amplification, primer dimer
	crAss004-Rev	AAGATGGCGTTAAAGGGGCA						
crAss 005	crAss005-For	CTGCTGCTCCTGAGAGTGTTA	494 to 596	3	Pass	Fail		False Positive, Spurious Band, Primer Dimer
	crAss005-Rev	TCCACTTGCGCAGAACTTTG						
crAss 006	crAss006-For	GCGCAAGTGGAAACCATTGA	586 to 703	3	Fail			False Positive
	crAss006-Rev	TGCCATTACGACCTGTACGTT						
crAss 007	crAss007-For	ACGTACAGGTCGTAATGGCA	684 to 811	3	Fail			Primer Dimer
	crAss007-Rev	ATGGCATCTGAACAGCACCA						
crAss 008	crAss008-For	GATGCCATTCAATCAGATACTGC	804 to 915	3	Fail			Primer Dimer
	crAss008-Rev	AGTCAGATACATACTAGCAAAGCCA						
crAss 009	crAss009-For	ATCTGACTGGTGTTGCTGTC	908 to 1057	3	Fail			Primer Dimer
	crAss009-Rev	CAGCAGGCTGTTCGATACCTAC						
crAss 010	crAss010-For	CGAACAGCCTGCTGACCCTA	1044 to 1189	3	Fail			Primer Dimer
	crAss010-Rev	CGTCATCACTCATTGCAGTAGC						
crAss 011	crAss011-For	GCCTAGAAAGAAAACTATGGCTGA	2733 to 2844	9	Fail			Primer Dimer
	crAss011-Rev	AACAGTTTCACTACTAACAACAGCA						
crAss 012	crAss012-For	GAGGGTAAACTTGTTGTTCCTG	2779 to 2906	9	Pass	Fail		False Positive, Spurious Band, Primer Dimer
	crAss012-Rev	GTTTGTTCTTCCTGTTTACGAACT						
crAss 013	crAss013-For	GATGCTAAGACTGATAAACGTGGT	2947 to 3059	9	Fail			Primer Dimer
	crAss013-Rev	AGAGAAGTATGACGTCTAAATGCAG						

Table B3 (continued)

crAss014	crAss014-For	TGCATTTAGACGTCATACTTCTCTT	3036 to 3170	9	Fail			Low amplification, primer dimer
	crAss014-Rev	CTACGAAGTCCACCATCAGCA						
crAss015	crAss015-For	AGAGTTACTGCTGATGGTGGA	3142 to 3251	9	Fail			Low amplification, primer dimer
	crAss015-Rev	ACTCGCTGACTATCACTAGCA						
crAss016	crAss016-For	TTCATGCAGAATGTCTAAGTCAAGA	3556 to 3648	9	Pass	Fail		False Positive, Primer dimer
	crAss016-Rev	AAACATCATTTTCAGGGTCAACA						
crAss017	crAss017-For	CTCCAGACCAATATGTTAGAGGT	3829 to 3963	11	Fail			Primer Dimer
	crAss017-Rev	GCCAATCAAGTTCAAATCCATCT						
crAss018	crAss018-For	TCAAACTAAGCTTGGTGTTTGTG	4754 to 4899	12	Fail			Low amplification
	crAss018-Rev	CACTAACACAACGACTAACACCTTT						
crAss019	crAss019-For	AACTCATACTAAAGGTGTTAGTCGT	4865 to 4974	12	Fail			No Product
	crAss019-Rev	TCTGTTCATTATTCTTACCATCACG						
crAss020	crAss020-For	GGCTATTGGAAAGATAAACTTGCT	5247 to 5410	12	Fail			No Product
	crAss020-Rev	AAATCAATATAAGAACGAGCAGCAG						
crAss021	crAss021-For	TTGTAGGTAAATCCAAAGAGAAAGG	6514 to 6660	15	Fail			Low amplification, primer dimer
	crAss021-Rev	TTCTGTAACTTTCACCAGTAATGC						
crAss022	crAss022-For	AAAGTTGTTTATGCTCGTCGATT	6672 to 6799	15	Fail			No Product
	crAss022-Rev	CCTTTAGCTATAAGTCCAGCAGC						
crAss023	crAss023-For	TTTATATTCCTTTTGCTGCTGGACT	6763 to 6872	15	Pass	Fail		Sensitivity
	crAss023-Rev	TTCACTATCACCACGATTATAACAC						
crAss024	crAss024-For	ACTAAAGACCAAGAAATAGCTGTGC	7013 to 7122	16	Fail			Low amplification, primer dimer
	crAss024-Rev	GTCTTACCTGTTCCTCCAGCA						
crAss025	crAss025-For	TCTTGCACAGCCTTGGGATA	7054 to 7193	16	Fail			Low amplification, primer dimer
	crAss025-Rev	GTGTAGGAGCAGCACAACCTA						
crAss026	crAss026-For	TATAGGTTGTGCTGCTCCTACAC	7171 to 7332	16	Fail			Primer Dimer
	crAss026-Rev	TCTTTTCCAACAGGATTAAAGGCA						
crAss027	crAss027-For	AGCTTCTATGCTTAATGCTAAACTT	7369 to 7479	16	Fail			Primer Dimer
	crAss027-Rev	TCATTAACAGGTGGAAGCTGACT						
crAss028	crAss028-For	TGACTCTAGTCAGCTTCCACC	7450 to 7548	16	Pass	Pass	Fail	Sensitivity
	crAss028-Rev	TCTCCTTGTCGTACAACTTCTTT						
crAss029	crAss029-For	AAGAAGTTGTACGACAAGGAGAT	7527 to 7625	16	Fail			Primer Dimer
	crAss029-Rev	AATCAAGAAACTTCCATCCATTCTT						

Table B3 (continued)

crAss030	crAss030-For	GGATTTTATGTTTGTGGTCAAACTG	7667 to 7798	16	Fail			Primer Dimer
	crAss030-Rev	ATTCCATTGTGCAACACGAC						
crAss031	crAss031-For	TTCAAGCTATTCATGGTGGTGCT	7992 to 8135	16	Fail			Low amplification, primer dimer
	crAss031-Rev	TACTTCCACGTTCAGAACTACTAGC						
crAss032	crAss032-For	AGCTAGTAGTTCTGAACGTGGA	8110 to 8243	16	Pass	Fail		Spurious band, sensitivity
	crAss032-Rev	TTGCAAAGCCATAATCTAAATCTCT						
crAss033	crAss033-For	TCTCATAGAGCACAAGGTTCTAC	8249 to 8428	16	Fail			Low amplification
	crAss033-Rev	TCGTCAACAGTTTTACTTGCC						
crAss034	crAss034-For	TGGCAAGTAAAACTGTTGACG	8407 to 8539	17	Fail			No Product
	crAss034-Rev	TCACCTCTAGGAAATAGGAAAACA						
crAss035	crAss035-For	TGGCTACTGGCGGTAAATATG	8803 to 8912	17	Fail			No Product
	crAss035-Rev	CCAATTAATATCATCAGCAAAACGC						
crAss036	crAss036-For	GGAACTATCCATTGTGTTCAACCTC	11349 to 11490	20	Fail			No Product
	crAss036-Rev	GACCATCAACAATTTCAGTTCTTCT						
crAss037	crAss037-For	AAAGATACTATTGATACTGCCAAGC	11661 to 11774	20	Fail			Low amplification
	crAss037-Rev	CTCTACGGGGAGTAAAGCCC						
crAss038	crAss038-For	CAGAGAGCCAATTTTGTTTGTTAAT	11780 to 11924	20	Fail			Low amplification
	crAss038-Rev	AGCACTACGTCCATCTTGAGT						
crAss039	crAss039-For	GTACTCAAGATGGACGTAGTGC	11902 to 12040	20	Fail			Primer Dimer
	crAss039-Rev	CGTTCAGCAAGAACAAGAGGAT						
crAss040	crAss040-For	GTTCTTGCTGAACGAGTTAAGACT	12027 to 12188	20	Pass	Fail		Sensitivity, Primer Dimer
	crAss040-Rev	AGTAGCAGCATATTCACCTCTCTT						
crAss041	crAss041-For	CAAGTCTGGAAGCAAGAAAAGACA	12293 to 12432	20	Fail			No Product
	crAss041-Rev	CGCTTGTTTCTAGGCTGTCA						
crAss042	crAss042-For	GAAACAAGCGTAGACGAATGGA	12423 to 12566	20	Fail			Low amplification
	crAss042-Rev	TGCTTCCTCAATAGTACCAGCTA						
crAss043	crAss043-For	TGTTCAACATTCTGTTGCTCCA	12792 to 12923	20	Fail			No Product
	crAss043-Rev	GCAAGACCATCTTCAACAGACTT						
crAss044	crAss044-For	AGTCTGTTGAAGATGGTCTTGCT	12902 to 13051	22	Fail			No Product
	crAss044-Rev	TCACGCTGCCTTTGACAAGT						
crAss045	crAss045-For	CTTGTCAAAGGCAGCGTGAG	13033 to 13172	22	Pass	Fail		False Positive, Primer dimer
	crAss045-Rev	GTTTGTGCCTCTTTGGGAGT						

Table B3 (continued)

crAss046	crAss046-For	TTACTCCCAAAGAGGCACAAACT	13151 to 13323	22	Fail			No Product
	crAss046-Rev	AGAAGCAATTTGTTTAGCTTCAGCA						
crAss047	crAss047-For	GCTTCTACTGGTAAAGTTCCTGTT	13318 to 13430	22	Fail			Primer Dimer
	crAss047-Rev	TCACCAGTAGCAACCATATTACG						
crAss048	crAss048-For	AGAACGTAATATGGTTGCTACTGGT	13404 to 13548	22	Fail			No Product
	crAss048-Rev	ATTAGCAATTTCACGAGCACC						
crAss049	crAss049-For	AAGATGCTATTGTTGCTGGATGT	13442 to 13579	22	Fail			Low amplification
	crAss049-Rev	CTTCTGTGCTTAAACAGCCCAT						
crAss050	crAss050-For	AGAAAGGTTTGAGTTTTGGTATGGG	13740 to 13854	23	Fail			Low amplification
	crAss050-Rev	ACTTGGGCAGTCATACGGAA						
crAss051	crAss051-For	GTTTCCGTATGACTGCCCAA	13833 to 13962	23	Fail			Primer Dimer
	crAss051-Rev	CAGAAAGCAACAACTTCGGGTA						
crAss052	crAss052-For	CCCGAAGTTGTTGCTTTCTG	13943 to 14091	23	Fail			Primer Dimer
	crAss052-Rev	ACACGAGTAGTACAAGGATTACCA						
crAss053	crAss053-For	ACTCGTGTTCGTATGACTAAGAATG	14084 to 14206	23	Fail			False Positive
	crAss053-Rev	ACCTTCACGACCAAGAGCAG						
crAss054	crAss054-For	TAAAGCTGCTCTTGGTCGTGAAG	14182 to 14322	23	Fail			Primer Dimer
	crAss054-Rev	GTTACACCAATACCAGACAAAGCA						
crAss055	crAss055-For	ATGTTTGGAAACAGTTGAAAGCTGA	14343 to 14452	23	Fail			Primer Dimer
	crAss055-Rev	ACCGTTATTAACAACAGCTTCTTGC						
crAss056	crAss056-For	GCTGAACAAACTGCTAATGCAGA	14712 to 14860	24	Pass	Pass	Pass	
	crAss056-Rev	TCAAGATGACCAATAAACAAGCCA						
crAss057	crAss057-For	GCCTTCATTTGCTGGTATGAGT	14909 to 15054	24	Fail			Low amplification
	crAss057-Rev	AAGCACCTTTACCACCGGG						
crAss058	crAss058-For	ATGTATTTGTTCTTAAAGGTCGTGA	15097 to 15212	24	Fail			Low amplification
	crAss058-Rev	AACTTCCTCAGCTTCAACAGGT						
crAss059	crAss059-For	AGCTGAGGAAGTTATAGCTGGTT	15200 to 15373	24	Fail			Low amplification, primer dimer
	crAss059-Rev	TTACTTCCAACAACAGCACGC						
crAss060	crAss060-For	TGGTAAATGGCGTGCTGTTG	15344 to 15469	24	Fail			Primer Dimer
	crAss060-Rev	GGGTCAATATGCAAACTAGGAGC						
crAss061	crAss061-For	AGAACAATCTGCTCCTAGTTTGC	15437 to 15574	24	Fail			Primer Dimer
	crAss061-Rev	GGAGCAGCAGTTTGAACACC						

Table B3 (continued)

crAss062	crAss062-For	TTGCTATGCCTGGTGTTGGTG	15526 to 15618	24	Fail			No Product
	crAss062-Rev	GACTAAAATCTCCACCGCCATT						
crAss063	crAss063-For	TACAACATTGCATAGATACAGGAGA	15896 to 16029	25	Fail			Primer Dimer
	crAss063-Rev	ATCTCCCCACCAATATCCAGC						
crAss064	crAss064-For	TGCTGCTGCAACTGTACTCT	16038 to 16177	25	Pass	Pass	Pass	
	crAss064-Rev	CGTTGTTTTCATCTTTATCTTGTCC						
crAss065	crAss065-For	CAACGATTATAATATTGCTAGGGCT	16173 to 16290	25	Fail			No Product
	crAss065-Rev	AACACCAAATTGTCCCCAATACT						
crAss066	crAss066-For	GGGGACAATTTGGTGTTAATCTT	16274 to 16423	25	Fail			No Product
	crAss066-Rev	GCCCAAGAACGTAACCATAACAT						
crAss067	crAss067-For	AATCATTCCATCCTCTACGGGG	16615 to 16742	25	Fail			Low amplification, primer dimer
	crAss067-Rev	TTCCGATTAAGTTTACTACGAAGCC						
crAss068	crAss068-For	TGGTGTTGTTAATATTCCACATGAA	16662 to 16798	25	Fail			Primer Dimer
	crAss068-Rev	TCACAGCTTCCATAAGACCAGT						
crAss069	crAss069-For	ACTTGGCGTTAAAGATTTTGCAG	16845 to 16985	25	Fail			No Product
	crAss069-Rev	TTACTTTCTTCCGTATCCCAAGT						
crAss070	crAss070-For	TGCGTCCAATTACTGAAGAGGA	17031 to 17205	26	Fail			Primer Dimer
	crAss070-Rev	CCACCATTATCAGTATCAAGTAGGC						
crAss071	crAss071-For	CCTACTTGATACTGATAATGGTGGA	17182 to 17331	26	Fail			False Positive
	crAss071-Rev	GCTGGCTTACCTATAAGACAATG						
crAss072	crAss072-For	GTCTTATAGGTAAGCCAGCTCGT	17313 to 17462	26	Fail			Primer Dimer
	crAss072-Rev	TCAATTGCATCAGCCATTGTAGA						
crAss073	crAss073-For	TGATAAATCTACAATGGCTGATGC	17433 to 17525	27	Fail			Primer Dimer
	crAss073-Rev	GTTTGTCTAGGACGAGAACCA						
crAss074	crAss074-For	ATCCTGATGGTTCTCGTCCT	17498 to 17618	27	Fail			No Product
	crAss074-Rev	TGAGCTTTTTCTGGATGATGTTCT						
crAss075	crAss075-For	CTAGAACATCATCCAGAAAAAGCTC	17593 to 17704	27	Fail			Primer Dimer
	crAss075-Rev	CTGGGTTTCTTCTCTCGAAATCAA						
crAss076	crAss076-For	CGAGAGAAGAAACCCAGTACGA	17688 to 17815	27	Fail			No Product
	crAss076-Rev	TTGCAATACCGGTGCTAACG						
crAss077	crAss077-For	AAGAACCAAGGAAACGCACT	17720 to 17858	27	Fail			Primer Dimer
	crAss077-Rev	GCTTTACGCTGAGCAATCGT						

Table B3 (continued)

crAss078	crAss078-For	ATTAACGTTAGCACCGGTATTGC	17791 to 17900	27	Fail			Primer Dimer
	crAss078-Rev	CCATTAAATGCAAAACTAACAGCTT						
crAss079	crAss079-For	GCTAGAGTTTGTATATGTTGTGGCA	20535 to 20649	32	Fail			No Product
	crAss079-Rev	ATTCTTTACATACAATAGCCTCAGC						
crAss080	crAss080-For	ACTCCACTTGGAAATTCTGTTGA	20574 to 20698	32	Fail			Primer Dimer
	crAss080-Rev	TCCATGTCACAAGCTATAAAGCAA						
crAss081	crAss081-For	TTGCTTTATAGCTTGTGACATGGA	20675 to 20795	32	Fail			No Product
	crAss081-Rev	AATAGGCTGTATAGAATCAAGCTCT						
crAss082	crAss082-For	CCTGAACTTGAAGAACGCCTTG	22187 to 22326	35	Fail			Primer Dimer
	crAss082-Rev	TTACGAAGAACAGCTAATCGTTGC						
crAss083	crAss083-For	GAGCAACGATTAGCTGTTCTTC	22301 to 22411	35	Pass	Fail		False Positive, Primer Dimer
	crAss083-Rev	ATTACCAGACTTACCTAAATCTCCA						
crAss084	crAss084-For	TTTAGTAGAGATGTGCTTCTTGAAC	22562 to 22671	35	Fail			Low amplification, primer dimer
	crAss084-Rev	TCATCAAACGTAAAATAACCTCCAG						
crAss085	crAss085-For	TGGAGGTTATTTTACGTTTGATGAT	22648 to 22797	35	Fail			Primer Dimer
	crAss085-Rev	ACACTGCTACTTCTAGTCATTTCA						
crAss086	crAss086-For	GGGCTTATTCTGGTGCTATTGATG	23111 to 23260	37	Fail			No Product
	crAss086-Rev	TTCGGTCAAGTTCTTCATCAGTTC						
crAss087	crAss087-For	GTGAACTAGTTGGTAAAATGCCTAT	23201 to 23330	37	Fail			Primer Dimer
	crAss087-Rev	AAGGCGTTATCACACTTACGA						
crAss088	crAss088-For	CGCCTTTGCTACCTATCGTACTG	23325 to 23474	37	Fail			Low amplification, primer dimer
	crAss088-Rev	CCCCAAAATCCAGCAGTTTGCC						
crAss089	crAss089-For	TGCTGGATTTTGGGGAAATGG	23460 to 23595	37	Fail			No Product
	crAss089-Rev	AATCTTAACTCCACCACTACCATCA						
crAss090	crAss090-For	CTCATGATGGTAGTGGTGGAGT	23567 to 23692	37	Fail			Primer Dimer
	crAss090-Rev	GAACACTAGTTGTATGTCGAAAACT						
crAss091	crAss091-For	TTCAACGTTTAAAAGATACAGGACA	23846 to 23994	37	Fail			No Product
	crAss091-Rev	ATCTCTTTGACCGGGACCAT						
crAss092	crAss092-For	TGATGGTCCCGGTCAAAGAG	23973 to 24084	37	Fail			Primer Dimer
	crAss092-Rev	ACTATCAAATCGCTTAGTACCTTCA						
crAss093	crAss093-For	TACAGCTTGGGGTGCAGTTA	24003 to 24138	37	Fail			Primer Dimer
	crAss093-Rev	TTCTGCAAGTGCAAAAGCAT						

Table B3 (continued)

crAss094	crAss094-For	TGGTTCAACCCATTCTACCTAC	24243 to 24392	38	Fail			Primer Dimer
	crAss094-Rev	GTTCACCATTAGCATCATTTCCAA						
crAss095	crAss095-For	TATTGGAAATGATGCTAATGGTGAA	24367 to 24476	38	Fail			Low amplification, primer dimer
	crAss095-Rev	TATAGTCTAAAGTACCAGGGGCA						
crAss096	crAss096-For	ATGGGCTATCGTGGAGAACT	25489 to 25607	40	Fail			Primer Dimer
	crAss096-Rev	GGTAAATGCATATGTTCTCTCAACT						
crAss097	crAss097-For	AGTTGAGAGAACATATGCATTTACC	25583 to 25709	40	Fail			No Product
	crAss097-Rev	CCATCACAATTATATGGCGGAGT						
crAss098	crAss098-For	AGTGCTGAACGTATTAGTTGGA	25741 to 25850	40	Fail			No Product
	crAss098-Rev	TAAGCTTTAGCTCCTCCACCG						
crAss099	crAss099-For	AGGTGAATGGATTCCTGCTGT	25933 to 26046	41	Fail			Spurious Bands
	crAss099-Rev	TCATTACCTTCAACTTTAGCATCAC						
crAss100	crAss100-For	AGTGATGCTAAAGTTGAAGGTAATG	26021 to 26150	41	Fail			No Product
	crAss100-Rev	GCAAACTCAGGTAGAGCAAGAG						
crAss101	crAss101-For	AGGCTTCATAAAGCTCTTGCTCT	26116 to 26234	42	Fail			No Product
	crAss101-Rev	GGAGCATCACTAATACCAGCACT						
crAss102	crAss102-For	TTGCTCTACCTGAGTTTGCC	26132 to 26280	42	Fail			No Product
	crAss102-Rev	ATCGCTTTGCACACCACG						
crAss103	crAss103-For	TGGTAGACAAGAAGCGCGTA	26193 to 26315	42	Fail			No Product
	crAss103-Rev	CCTAGCAGCTTGCGAAGATA						
crAss104	crAss104-For	AAGGCTGTTGCTGGTAAGAAT	26608 to 26757	44	Fail			No Product
	crAss104-Rev	TTGTCTTGTAGGTTCATCTCCAG						
crAss105	crAss105-For	CTGGAGATGAACCTACAAGACAAA	26735 to 26847	44	Fail			No Product
	crAss105-Rev	TTTCCACATCCAAGCAATAGCATC						
crAss106	crAss106-For	GCTTTAATAGCAGCAGCTTTACGTT	31336 to 31425	52	Fail			No Product
	crAss106-Rev	TACTCAACAGTTGCCTACTATACCT						
crAss107	crAss107-For	AGGTATAGTAGGCAACTGTTGAG	31401 to 31538	52	Fail			No Product
	crAss107-Rev	GGTTGGACAGCCAATGTTCC						
crAss108	crAss108-For	GTATTATTAGGAACATTGGCTGTCC	31510 to 31620	52	Fail			No Product
	crAss108-Rev	TGATATGGGTTATCTTCCAGCTTT						
crAss109	crAss109-For	GCTGGAAGATAACCCATATCAACAT	31600 to 31749	52	Fail			No Product
	crAss109-Rev	GTACAAAGTGGGAACAGGATATGAT						

Table B3 (continued)

crAss110	crAss110-For	TATCCTGTTCCCACTTTGTACTT	31729 to 31851	52	Fail			No Product
	crAss110-Rev	TGGTTCAGCTGTTGTTGGTAA						
crAss111	crAss111-For	GCAGCAGTTCTATCAATATCAGTC	32050 to 32163	52	Fail			No Product
	crAss111-Rev	TAGTGAACAAGTTGAAGAAAATGGA						
crAss112	crAss112-For	CACGTTTTGCTTTAGCTTTATCAGT	32282 to 32371	52	Fail			No Product
	crAss112-Rev	AAGATTTAGAACAACTCAAGACTGG						
crAss113	crAss113-For	CAAAACCAGTCTTGAGTTGTTCTAA	32342 to 32472	52	Fail			No Product
	crAss113-Rev	GAGCGAAGAAGAAATTAATGTAGGA						
crAss114	crAss114-For	CCTACATTAATTTCTTCTTCGCTCT	32449 to 32574	52	Fail			No Product
	crAss114-Rev	AGGCAATGGTATGCCTTATGTTG						
crAss115	crAss115-For	GAACGAATAATACCAACATAAGGCA	32539 to 32652	52	Fail			No Product
	crAss115-Rev	AGTTGATGGACGTAAAATACCTGA						
crAss116	crAss116-For	TGCACGTTCTTCAGGTATTTTACG	32619 to 32752	52	Fail			No Product
	crAss116-Rev	ATTTAAATTGGGCGTTTGAAGAAC						
crAss117	crAss117-For	CTTTAAGTTCTTCAAACGCCCA	32723 to 32834	52	Fail			No Product
	crAss117-Rev	ATGAAAGTTCCTGAATATCGTGAAG						
crAss118	crAss118-For	ATTACCATTAGCATCACGTTGACT	32859 to 33008	52	Fail			No Product
	crAss118-Rev	GCTGAAAGACTTCGTGCTTATCAA						
crAss119	crAss119-For	TTCAGCTACATTCAAATCATGACCA	33003 to 33112	52	Fail			No Product
	crAss119-Rev	AAAGTCAAGATGAAGAACTTGGACG						
crAss120	crAss120-For	CGTCCAAGTTCTTCATCTTGACT	33088 to 33211	52	Fail			No Product
	crAss120-Rev	AAGGAGATGATGGTATTCCTAAAGA						
crAss121	crAss121-For	TCTTTAGGAATACCATCATCTCCTT	33187 to 33308	52	Fail			No Product
	crAss121-Rev	AAACCTATCACAGTTCAAGATGAAG						
crAss122	crAss122-For	TCTTCATCTTTTCTTCCCAAACAGG	33425 to 33534	52	Fail			No Product
	crAss122-Rev	TAAATCCAGAGCTGCTAAAGATGGT						
crAss123	crAss123-For	ACCATCTTTAGCAGCTCTGGA	33510 to 33653	52	Fail			No Product
	crAss123-Rev	CGTGCTAATCGTAACTTTGTATTGC						
crAss124	crAss124-For	CGACCTTCTTCAACAGTACCCA	33586 to 33698	52	Fail			No Product
	crAss124-Rev	GCTTATGGAGATACTGGATTTATGG						
crAss125	crAss125-For	AGCATCAAATTCTTTCATAAGACCA	33696 to 33788	52	Fail			No Product
	crAss125-Rev	CATTCTGCTAGAAAAGATTGGTTTG						

Table B3 (continued)

crAss126	crAss126-For	AAGCATTAAGAACATTCTGGGC	34157 to 34266	52	Fail			No Product
	crAss126-Rev	TAGTGTTGATGACCAAACTCCTAT						
crAss127	crAss127-For	GCTATTCCGCCCTCACTCAA	35020 to 35168	52	Fail			No Product
	crAss127-Rev	TGGAATCATCCCAATCCTCTTATT						
crAss128	crAss128-For	GGGATGATTCCAAGCTCTATCAAA	35157 to 35271	52	Fail			No Product
	crAss128-Rev	TAATGACGAAAGAAATAATCGTCGT						
crAss129	crAss129-For	ACATTACGACGATTATTTCTTTCGT	35242 to 35361	52	Fail			No Product
	crAss129-Rev	TGCTAGACTTACTCCTGCTCAA						
crAss130	crAss130-For	TTTGAGCAGGAGTAAGTCTAGCA	35339 to 35467	52	Fail			No Product
	crAss130-Rev	GAGCTTTAGAAGTTGGACGAGT						
crAss131	crAss131-For	ACTAATGTACCACTTACTCGTCCAA	35431 to 35540	52	Fail			No Product
	crAss131-Rev	TTACCTGTTACTGTTGTTAATGGTG						
crAss132	crAss132-For	TCACCATTAACAACAGTAACAGGT	35515 to 35649	52	Fail			No Product
	crAss132-Rev	AGCAGAATGGACTAAAGGTGGA						
crAss133	crAss133-For	TCCACCTTTAGTCCATTCTGCT	35628 to 35758	52	Fail			No Product
	crAss133-Rev	ATCCTAGTTTAGCTGCATTTTTACA						
crAss134	crAss134-For	TTGTAAAAATGCAGCTAAACTAGGA	35733 to 35867	52	Fail			No Product
	crAss134-Rev	TATGCTCCATCCTCTACGGG						
crAss135	crAss135-For	CCCCGTAGAGGATGGAGCAT	35847 to 35961	52	Fail			Spurious Bands
	crAss135-Rev	TCTTAGTGCTGATAAATATGGTGCT						
crAss136	crAss136-For	TTGTTTAGCACCATATTTATCAGCA	35931 to 36075	52	Fail			No Product
	crAss136-Rev	TCGTCTTAGAGGGGAAATGAATGG						
crAss137	crAss137-For	TAAAGCCTGCTGAACGCCA	36036 to 36145	52	Fail			Spurious Bands
	crAss137-Rev	TATTCAATTTAAGCGTTTCTGATGC						
crAss138	crAss138-For	TCAATTAGCTTTTTACGTCCAGCA	36184 to 36295	52	Fail			No Product
	crAss138-Rev	GTGTTCTTGCGGAAGGTACT						
crAss139	crAss139-For	ACCAAAACCAAGATTATGAGCAAC	36222 to 36353	52	Fail			No Product
	crAss139-Rev	GGATTTATGTGGCAACCTGCTA						
crAss140	crAss140-For	ATCAATAGCAGGTTGCCACA	36327 to 36453	52	Fail			Spurious Bands
	crAss140-Rev	CGATGTTATGAAAGAGGGGGCTA						
crAss141	crAss141-For	GCCCCCTCTTTCATAACATCG	36433 to 36542	52	Fail			No Product
	crAss141-Rev	CATAAGAATTTTGGTTGGTCTGGA						

Table B3 (continued)

crAss142	crAss142-For	CTTATCATCTCCAGACCAACCA	36510 to 36623	52	Fail			No Product
	crAss142-Rev	GTGATTAGTGAGCAAGAAGCTG						
crAss143	crAss143-For	AGCTTCTTGCTCACTAATCACA	36603 to 36751	52	Fail			No Product
	crAss143-Rev	GGTTTGATGCTGCTACTTCCG						
crAss144	crAss144-For	AAGTAGCAGCATCAAACCAATCA	36734 to 36882	52	Fail			No Product
	crAss144-Rev	TGCTGCATTTGAAGAAAATACTACT						
crAss145	crAss145-For	GTAGTATTTTCTTCAAATGCAGCAG	36859 to 36968	52	Fail			No Product
	crAss145-Rev	TTAAAACTCCCCGTGGAGGATAC						
crAss146	crAss146-For	GACTACTTGTATCCTCCACGG	36938 to 37049	52	Fail			No Product
	crAss146-Rev	ACTCTTGAAGGTTATAATGGCAATC						
crAss147	crAss147-For	GCCATTATAACCTTCAAGAGTAGCA	37029 to 37138	52	Fail			No Product
	crAss147-Rev	AAGCTAGTCGTAAGATTCTTGGT						
crAss148	crAss148-For	AAGTCCTACCAAGAATCTTACGAC	37109 to 37218	52	Fail			No Product
	crAss148-Rev	TATTCCAACTGAAATTACTCACGCT						
crAss149	crAss149-For	TCTAGCGTGAGTAATTTCAGTTGG	37191 to 37312	52	Fail			No Product
	crAss149-Rev	TTGTTGTTCCTGATGAATGGGT						
crAss150	crAss150-For	CAGGAACAACAACAGTACTACCA	37301 to 37432	52	Fail			No Product
	crAss150-Rev	TTGAACAAGTTCCTGATAGCTTACG						
crAss151	crAss151-For	ACGTAAGCTATCAGGAACTTGT	37407 to 37516	52	Fail			No Product
	crAss151-Rev	AGATTGGAGAGGTTGATGGCA						
crAss152	crAss152-For	GGAGTGCCATCAACCTCTCC	37492 to 37602	52	Fail			No Product
	crAss152-Rev	TAGTGGATGGCATGCTGCTC						
crAss153	crAss153-For	GCAGCATGCCATCCACTAATA	37585 to 37734	52	Fail			No Product
	crAss153-Rev	AGCTCTTCTTGAGTTTTTCGATTTA						
crAss154	crAss154-For	TTTTCAACTCCTTGTTGTTGAGCAT	37735 to 37873	52	Fail			No Product
	crAss154-Rev	AAACTTGTGCTGAACTTGGTATT						
crAss155	crAss155-For	CCAATACCAAGTTCAGCACAAGT	37849 to 37966	52	Fail			No Product
	crAss155-Rev	TGCTTGATAATCTTCCTAATCGAAC						
crAss156	crAss156-For	TGTTGAGGAACTTCTTGCTGAC	38014 to 38159	52	Fail			No Product
	crAss156-Rev	GTTGCACAACATAATCGAATGACA						
crAss157	crAss157-For	AGTGTCATTCGATTATGTTGTGC	38134 to 38274	52	Fail			No Product
	crAss157-Rev	AAATGCTGAGTTTGTTCTTATTCCT						

Table B3 (continued)

crAss158	crAss158-For	TGTCTAGAAACTTCAATACCAACAG	38281 to 38425	52	Fail			No Product
	crAss158-Rev	ATGCTGGACTTATTCAAAGTCTTAC						
crAss159	crAss159-For	CTGTAAGACTTTGAATAAGTCCAGC	38399 to 38513	52	Fail			No Product
	crAss159-Rev	GGCGGTGTTAATACTAATGATGCT						
crAss160	crAss160-For	TAACACCGCCTTTAAATGGTTC	38504 to 38593	52	Fail			No Product
	crAss160-Rev	TTATTGATAGACTCGATGACCAACT						
crAss161	crAss161-For	AAGTTGGTCATCGAGTCTATCAAT	38568 to 38716	52	Fail			No Product
	crAss161-Rev	AAGGTAATGCTGTTACTGTTCCT						
crAss162	crAss162-For	AACAGTATAAGGAACAGTAACAGCA	38685 to 38801	52	Fail			Spurious Bands
	crAss162-Rev	CAAGCTAGTGGTAATCCTTTTGGT						
crAss163	crAss163-For	GGATTACCACTAGCTTGTACTTCTT	38785 to 38875	52	Fail			No Product
	crAss163-Rev	TGTACGATATGTATGGTGGTGAC						
crAss164	crAss164-For	AGCAACACGTTGTTCTCCATTA	39027 to 39117	52	Fail			No Product
	crAss164-Rev	TAAGACTATTGACATCCTCTACGG						
crAss165	crAss165-For	GCTAAGTCCACGTGTAGACCC	39072 to 39219	52	Fail			No Product
	crAss165-Rev	TGGACACAAAGTTTTAGCTGGT						
crAss166	crAss166-For	CATTACATTAACAGCTTGAGCCA	39357 to 39483	52	Fail			No Product
	crAss166-Rev	TGCTGGATTGCGTAGTCAAC						
crAss167	crAss167-For	GACTACGCAATCCAGCAAGAC	39467 to 39639	52	Fail			No Product
	crAss167-Rev	ATATCGTAGTATGAGTGATGGTGAT						
crAss168	crAss168-For	TAATCACCATCACTCATACTACGAT	39613 to 39742	52	Fail			No Product
	crAss168-Rev	GAGTTGGAAATGGAACTCCTAAACT						
crAss169	crAss169-For	AGGAGTTCCATTTCCAACTCTTCT	39723 to 39867	52	Fail			No Product
	crAss169-Rev	GCAAAACGCTATGGTTGAAGC						
crAss170	crAss170-For	GCTTCAACCATAGCGTTTTGC	39847 to 39975	52	Fail			No Product
	crAss170-Rev	ACTTAATTCTCGTAATCCAGAAGGT						
crAss171	crAss171-For	TCAGGCATTTTAGTTCCTTTATCCT	40057 to 40185	52	Fail			No Product
	crAss171-Rev	TGCTAAGATTGGTCGTGTAGC						
crAss172	crAss172-For	CTACACGACCAATCTTAGCAATAG	40166 to 40315	52	Fail			No Product
	crAss172-Rev	TTCCTGCTGCTCGTGAAAAG						
crAss173	crAss173-For	ACTTTTCACGAGCAGCAGGA	40295 to 40406	52	Fail			Spurious Bands
	crAss173-Rev	GCTCGTGTAACTAATCCTAATACTG						

Table B3 (continued)

crAss174	crAss174-For	ACATTACTAGTATCAGCACTAGCAT	40600 to 40731	52	Fail			No Product
	crAss174-Rev	ACTTACTGCTGCTTCTCAACTT						
crAss175	crAss175-For	GAGAAGCAGCAGTAAGTTTAGGTA	40715 to 40894	52	Fail			Spurious Bands
	crAss175-Rev	TAGGCGGTTATCTTGATGAAACT						
crAss176	crAss176-For	AGTTTCATCAAGATAACCGCCTA	40872 to 41017	52	Fail			Spurious Bands
	crAss176-Rev	ATTATGGTGCATTACTTCGTTCTTT						
crAss177	crAss177-For	GAAGAAATGCTCCAAGACCGT	41045 to 41167	52	Fail			No Product
	crAss177-Rev	ATGCTGGACTTAAACGCGCTA						
crAss178	crAss178-For	TAGCGCGTTTAAGTCCAGCA	41147 to 41322	52	Fail			No Product
	crAss178-Rev	AGAAGAACGTGGTTTGTTTAGTACA						
crAss179	crAss179-For	TCAGCAATAGAAGCTACATCATCCA	41326 to 41466	52	Fail			No Product
	crAss179-Rev	GGTTGAAGACCCAATCTTTAAGAAG						
crAss180	crAss180-For	CTACATACAGCTAAAACTGATGCGT	41701 to 41790	53	Fail			No Product
	crAss180-Rev	GCTGAATTAACTAGCCGTTTACCA						
crAss181	crAss181-For	AGTTAATTCAGCCATATTTGACACT	41779 to 41920	53	Fail			No Product
	crAss181-Rev	AACTGCTAGTACAGAACAAACTGAA						
crAss182	crAss182-For	TCTGTACTAGCAGTTTCTTCTTCA	41906 to 42020	53	Fail			No Product
	crAss182-Rev	GTTTATTAGCTGCTGCTAGAGATG						
crAss183	crAss183-For	CATCTCTAGCAGCAGCTAATAAAC	41997 to 42144	53	Fail			No Product
	crAss183-Rev	GCTCTTATGACTGCACTTAATGAAA						
crAss184	crAss184-For	TTCATTAAGTGCAGTCATAAGAGC	42121 to 42257	53	Fail			No Product
	crAss184-Rev	GGCTTGATATTCAAGATGATGCTAC						
crAss185	crAss185-For	ATCATCTTGAATATCAAGCCAAGC	42238 to 42387	53	Fail			No Product
	crAss185-Rev	GCTGGTCAAAGAGTTTATCGAGT						
crAss186	crAss186-For	CTGTATCACGATGATAAACAGCATT	42417 to 42557	53	Fail			No Product
	crAss186-Rev	ATTTTGCTAATGACCAAACTGCTA						
crAss187	crAss187-For	TCAATAGCAGTTTGGTCATTAGC	42530 to 42679	53	Fail			No Product
	crAss187-Rev	GGATGGAACAATACCTGAACAGC						
crAss188	crAss188-For	GATAGCTGTTCAGGTATTGTTCCA	42653 to 42763	53	Fail			No Product
	crAss188-Rev	TAATGAAACTGCTGGTAATTGGC						
crAss189	crAss189-For	AACAGGACGAAGAGCATTATCTA	42976 to 43153	53	Fail			No Product
	crAss189-Rev	AGGTGCACATTTTAATGGTAAAGA						

Table B3 (continued)

crAss190	crAss190-For	CTCCAAGATTTTGTTTAAGTTCACC	43245 to 43355	53	Fail			No Product
	crAss190-Rev	AAGGTGATGCTCGACGTATT						
crAss191	crAss191-For	AATACGTCGAGCATCACCTT	43336 to 43457	53	Fail			Spurious Bands
	crAss191-Rev	ATCTTACTACTAATGGTATGCCTGT						
crAss192	crAss192-For	TACTTGACCTTGTACAACTACACCA	43495 to 43587	53	Fail			No Product
	crAss192-Rev	GATGCTAATGGCAAATGGAATGATG						
crAss193	crAss193-For	TCCATTTGCCATTAGCATCATCAA	43569 to 43723	53	Fail			No Product
	crAss193-Rev	TCATGAAGCTAGTATTAATCGTTGG						
crAss194	crAss194-For	GCCCAGCTAAATCAAAACTTGC	43773 to 43894	53	Fail			False Positive
	crAss194-Rev	TGCTGTTAAAGAAGAAGTTACACAA						
crAss195	crAss195-For	AGTCTTTCAAGACGTTCTTTAGCA	44228 to 44347	53	Fail			Primer Dimer
	crAss195-Rev	CGCCAGCACTAATAATCAAGGT						
crAss196	crAss196-For	AGTGCTGGCGTAAAGATAGTT	44338 to 44513	53	Fail			Primer Dimer
	crAss196-Rev	TGGATGATTTTGCTAATAGTGTTGA						
crAss197	crAss197-For	ACGTTCATTACCTGTAACAGTAGCA	44530 to 44639	53	Pass	Fail		False Positive, Primer Dimer
	crAss197-Rev	ATATTATTGACGGTCTTACAGGAGG						
crAss198	crAss198-For	TTTGTTCAAATCCAGCTTGACTAAG	44664 to 44813	53	Fail			Low amplification
	crAss198-Rev	ATTTCGAACGTGGTCAAATAGGT						
crAss199	crAss199-For	AATCAGTACCTATTTGACCACGTT	44784 to 44897	53	Fail			Low amplification, primer dimer
	crAss199-Rev	ATCAATCTCAAACTGAACAACCAAT						
crAss200	crAss200-For	AGTAGGCTCACTAGGAACCGC	44911 to 45037	53	Fail			Primer Dimer
	crAss200-Rev	AGAAGTTGCTCCTGTTAATCCTGA						
crAss201	crAss201-For	CCCCGTGGAGGATGGAGATG	44953 to 45083	53	Fail			No Product
	crAss201-Rev	AGATTGCTGAGATTGAAAACGATGT						
crAss202	crAss202-For	TTAACATCGTTTTCAATCTCAGCA	45056 to 45165	53	Fail			Primer Dimer
	crAss202-Rev	ATTATTCGAGATGCTTATGCTGCAC						
crAss203	crAss203-For	GCATTACGTAAATCGTCCATAGC	45245 to 45403	53	Fail			Primer Dimer
	crAss203-Rev	AGCAGAAGTTGTTAATGCTGAGAT						
crAss204	crAss204-For	ATCTCAGCATTAACAACTTCTGCT	45380 to 45534	53	Fail			Primer Dimer
	crAss204-Rev	CGTATTGCTCAATTAAATGAACGTC						
crAss205	crAss205-For	TTCAAATCCGCCACCAAGTC	45772 to 45899	53	Fail			Low amplification, primer dimer
	crAss205-Rev	ATAAACTTGGATTACAACAGCAAGA						

Table B3 (continued)

crAss206	crAss206-For	AAGTTCATTACTATCTTGCTGTTGT	45862 to 45972	53	Fail			Low amplification, primer dimer
	crAss206-Rev	TCTATGAACGCTGGTAATTTAGGAC						
crAss207	crAss207-For	ACGAAGTTCTTCTTCAGTACCAG	46015 to 46149	53	Fail			Primer Dimer
	crAss207-Rev	ACTGCTACTTTTCAAGCATATCAAG						
crAss208	crAss208-For	TATGCTTGAAAAGTAGCAGTACGAC	46130 to 46276	53	Fail			Primer Dimer
	crAss208-Rev	TTGGGGTGCTCTTGGTGGT						
crAss209	crAss209-For	CCCAAAATGCTTGCTCCCAT	46272 to 46418	53	Fail			Primer Dimer
	crAss209-Rev	GGACAGAAGGAGTTGAAGAAGC						
crAss210	crAss210-For	TGTCCATTCAGCTCTAACACCA	46414 to 46524	53	Fail			Primer Dimer
	crAss210-Rev	GCTGCTGCAATTACTAACGCTT						
crAss211	crAss211-For	AAGCGTTAGTAATTGCAGCAG	46503 to 46648	53	Pass	Fail		Primer Dimer
	crAss211-Rev	TGTTCTTTTTGACGTATGGCAAGTT						
crAss212	crAss212-For	GGAACTCCAATAGCAGTAGCTT	46859 to 47007	53	Fail			Primer Dimer
	crAss212-Rev	AGTACTGGTCTTGCTAAAGGTGT						
crAss213	crAss213-For	AAAGAAACACCTTTAGCAAGACC	46979 to 47095	53	Fail			Primer Dimer
	crAss213-Rev	TCCTAATGCTGCTTTTGATATTGGA						
crAss214	crAss214-For	AAAAGCAGCATTAGGATTTTCTCT	47080 to 47229	53	Fail			Primer Dimer
	crAss214-Rev	GGTTTTGCAGATTTAGCTGATGC						
crAss215	crAss215-For	CTGCAAAACCAATAGCTGTACCAAC	47220 to 47355	53	Fail			Primer Dimer
	crAss215-Rev	TATGTTAATCCTGTTGATGACCCAG						
crAss216	crAss216-For	TTAACAGCACGTTCTTTATCAAGTT	47303 to 47447	53	Fail			Spurious Bands
	crAss216-Rev	ACCGGGGTCGTGGTCAT						
crAss217	crAss217-For	CACGACCCCGGTCACTAATA	47436 to 47550	53	Pass	Fail		False Positive
	crAss217-Rev	GGTGGTAATAAAACACCTAATCCTG						
crAss218	crAss218-For	CTATGACCGCCACAGCGATA	52841 to 52950	56	Pass	Fail		False Positive
	crAss218-Rev	ACTGGTCTTGAACGTCCTGA						
crAss219	crAss219-For	TCAGGACGTTCAAGACCAGTT	52931 to 53073	56	Fail			False Positive
	crAss219-Rev	AAAGCATGGATTGGTGCTGC						
crAss220	crAss220-For	CTTTTCTACGTCCGCCACAT	53070 to 53209	56	Fail			Primer Dimer
	crAss220-Rev	TCAAGAAGCGGAACAGCGT						
crAss221	crAss221-For	TGTTTACGCTGTTCCGCTTC	53186 to 53335	56	Fail			Primer Dimer
	crAss221-Rev	TGCTCTTCCCGGTTTAGGTAG						

Table B3 (continued)

crAss222	crAss222-For	ACCAGCTTTACGTCGAACTCT	53356 to 53501	56	Fail			Primer Dimer
	crAss222-Rev	CTGAAGCTCAAGCTGCACAAG						
crAss223	crAss223-For	ATCTTTATGGTCTTGTGCAGCTTG	53470 to 53579	56	Fail			Spurious Bands
	crAss223-Rev	TTATTGGTGCAGCTGCTAGTATTG						
crAss224	crAss224-For	AGAACTCATTCCAACCACTATCA	58165 to 58300	59	Fail			No Product
	crAss224-Rev	CATCCTGTGTTTAATCTTTCTGCTA						
crAss225	crAss225-For	TAGCAGAAAGATTAAACACAGGATG	58276 to 58390	59	Fail			Low amplification, primer dimer
	crAss225-Rev	TGTGAGATTCCTGTTATACCTGCTA						
crAss226	crAss226-For	CCATTAGCAGGTATAACAGGAATCT	58362 to 58524	59	Fail			False Positive
	crAss226-Rev	CCCGTGGAGGATATGATAAGTCAA						
crAss227	crAss227-For	GGGGTCTGACAATGAGAATGAAC	58524 to 58671	59	Fail			Primer Dimer
	crAss227-Rev	TGCCCGATGATTGTTGTCCTAA						
crAss228	crAss228-For	TTAGTAGCACGCTCCGGACT	58689 to 58801	59	Fail			No Product
	crAss228-Rev	GGTTGTGAAGTACCTAGAGCTGA						
crAss229	crAss229-For	AGACCGTCATCAATAAGACAACATT	58821 to 58942	59	Fail			Primer Dimer
	crAss229-Rev	CCTGCTTGGTGTGTTGATGC						
crAss230	crAss230-For	TTACCATTATTATGTGCTCCACGTT	59062 to 59151	59	Fail			Low amplification, primer dimer
	crAss230-Rev	TTATCTTGAACGAGAACCTAATGCT						
crAss231	crAss231-For	TACCATCATACTTAAGTCCACGAA	59363 to 59477	60	Pass	Fail		False Positive, Primer Dimer
	crAss231-Rev	GTTGATTGGAACGCTACTAGACT						
crAss232	crAss232-For	TAAGTCTAGTAGCGTTCCAATCAA	59453 to 59582	60	Fail			Primer Dimer
	crAss232-Rev	GTTCATAAATGTGTTCTTGAAGGC						
crAss233	crAss233-For	ATAAGCGTAGCCTTCAAGAACAC	59550 to 59667	60	Fail			Primer Dimer
	crAss233-Rev	TGTTGCTCTTAAACTCGCTACAA						
crAss234	crAss234-For	TCAGAACGTTTTGTAGCGAGT	59635 to 59750	60	Fail			Low amplification, primer dimer
	crAss234-Rev	AGTTATACAGTTGGAGAAAGACGTA						
crAss235	crAss235-For	TACGTCTTTCTCCAACTGTATAACT	59726 to 59901	60	Fail			Primer Dimer
	crAss235-Rev	GATGAAAGCTGATTGTCGTGCTT						
crAss236	crAss236-For	GGAGTTTCATTTGTTTGACGAGGA	60031 to 60162	62	Fail			Primer Dimer
	crAss236-Rev	TGACCCTATTGATGATGAAGCTAA						
crAss237	crAss237-For	CAGTTTCGGTAGGAACAAGATAAGG	60188 to 60336	62	Fail			Low amplification, primer dimer
	crAss237-Rev	TCTTGCAGGTATGTGTAATCTTCCT						

Table B3 (continued)

crAss238	crAss238-For	ACAGGAAGATTACACATACCTGC	60310 to 60455	62	Pass	Fail		Primer Dimer
	crAss238-Rev	GAAGTTCCAAAGCCAGTTAGATT						
crAss239	crAss239-For	ACTGGCTTTGGAACTTCTTGT	60439 to 60567	62	Pass	Fail		Spurious Band, sensitivity
	crAss239-Rev	TGTTGATAAAGGTTTGCAACAACG						
crAss240	crAss240-For	TCACCTATTGAAGTCTTACCATCA	66639 to 66744	72	Fail			Primer Dimer
	crAss240-Rev	TTGGTTTTATTGCAGGTAATTCTGA						
crAss241	crAss241-For	GTTGGCTTATCAAGAGTTAAAGGAG	66855 to 66964	72	Fail			Low amplification, primer dimer
	crAss241-Rev	ATTCATCGTGATGGTTATTGTTGGT						
crAss242	crAss242-For	AACCACCACCATGACGAATTA	66967 to 67088	72	Fail			Primer Dimer
	crAss242-Rev	TGGTGAAGTTATTAAAAGAGTTGCT						
crAss243	crAss243-For	TCTTTATAACCACCATGTCCATCA	67098 to 67207	72	Fail			Primer Dimer
	crAss243-Rev	GCTTGGCAATTAAAGAATCATGAAA						
crAss244	crAss244-For	TCATGATTCTTTAATTGCCAAGCAT	67185 to 67294	72	Fail			Primer Dimer
	crAss244-Rev	AATCTTTGTGCTAATTGGTATGGAA						
crAss245	crAss245-For	ACATGACCTCTACTCAAACTTCCA	67251 to 67372	72	Fail			Low amplification, primer dimer
	crAss245-Rev	TCAGGTGCTGAATGTGTTATGT						
crAss246	crAss246-For	CACCTGAATTAGTATTATCGCCAC	67366 to 67515	72	Fail			Low amplification, primer dimer
	crAss246-Rev	GGGGTAGTGATGATGGTCAAG						
crAss247	crAss247-For	ACCCCAAATGTATCTAAGTTGGAC	67511 to 67660	72	Fail			Primer Dimer
	crAss247-Rev	CGTCAAGGAGGAATTGCTCAAC						
crAss248	crAss248-For	GTCATTCCAGCAACTGGGGTA	67700 to 67825	73	Fail			Primer Dimer
	crAss248-Rev	AGCTATTAATGACGGATGTGGAG						
crAss249	crAss249-For	GAAGAAGTATCTCCCATTACTCTCC	68015 to 68159	73	Pass	Fail		Primer Dimer
	crAss249-Rev	AACTTACAGGTATGGTTGGTGAAG						
crAss250	crAss250-For	CTGACCTTTAGCCCAATTACTACA	68086 to 68235	73	Fail			Primer Dimer
	crAss250-Rev	ACAATGCCTGTTGGTTGTGA						
crAss251	crAss251-For	ACCATATCTTGAGTAAGAACAGCTT	68297 to 68440	73	Fail			Low amplification, primer dimer
	crAss251-Rev	TCTTTCTGATGAACGTCCAAGACT						
crAss252	crAss252-For	TGGACGTTCATCAGAAAGACAAG	68422 to 68541	73	Fail			Primer Dimer
	crAss252-Rev	AGATGGACTATTGTAGAACCTACTG						
crAss253	crAss253-For	TTCCATAAGCAGGAGTGTTGAC	68472 to 68581	73	Pass	Fail		False Positive, Spurious Band
	crAss253-Rev	GACTATTCATGATGGAGTTGGATTT						

Table B3 (continued)

crAss254	crAss254-For	TCAGTAGGTTCTACAATAGTCCATC	68516 to 68651	73	Fail			Low amplification, primer dimer
	crAss254-Rev	GAACAGATTTAGGTAATACTGCTGG						
crAss255	crAss255-For	CCACGGGGAGTAAGACCTTG	68763 to 68872	73	Fail			Primer Dimer
	crAss255-Rev	TCCAGGTTGTTGTAATGTTAATCGT						
crAss256	crAss256-For	GATTAACATTACAACAACCTGGACT	68850 to 68997	73	Pass	Fail		Primer Dimer
	crAss256-Rev	CCAATTGCTGATGGAACCATTGAT						
crAss257	crAss257-For	CCAATTAGTAGAACGATTAATGCCA	76436 to 76561	77	Fail			Low amplification
	crAss257-Rev	GTTATAGCACCTAGTAACTGGAGA						
crAss258	crAss258-For	ACTTGATGGTTATTATCACCGGATT	76560 to 76678	77	Fail			Low amplification
	crAss258-Rev	GCTGCTCCGGGTCGTAATG						
crAss259	crAss259-For	TGAAGAGGTAAATCAGTACCACC	76614 to 76728	77	Fail			Low amplification
	crAss259-Rev	ATCTTGATTGGAAAACAGTTCTTGA						
crAss260	crAss260-For	CCCATAACAAAACGTCCTTGTGC	76776 to 76887	77	Fail			Low amplification, primer dimer
	crAss260-Rev	AAGCTTGGATTGCTGCCGTT						
crAss261	crAss261-For	TTCCCAACCATTAGGAGGAGTT	76811 to 76936	77	Fail			Primer Dimer
	crAss261-Rev	CAAGGTGGTGGAAGTGGACAA						
crAss262	crAss262-For	CCACCTTGTCCAAATCCTTCTCT	76929 to 77074	77	Fail			Primer Dimer
	crAss262-Rev	CTTGGTGAAGCTAGAACAGGTAAGT						
crAss263	crAss263-For	GCAAGAAGCCAAGCATAGAACT	77622 to 77745	77	Pass	Fail		False Positive, Spurious Band
	crAss263-Rev	GGGATGGTGAAAAATGGGTAAG						
crAss264	crAss264-For	TACTTACCCATTTTTCACCATCCC	77722 to 77861	77	Fail			Low amplification, primer dimer
	crAss264-Rev	TGGAAATTGGGCTGTTGGT						
crAss265	crAss265-For	ACAGCCCAATTTCCATCAAGAG	77847 to 77977	77	Fail			No Product
	crAss265-Rev	ATTATTCTTTGGAGAGGTGGTAGAA						
crAss266	crAss266-For	AGTTTGTCCAGCTAAACAATAATCA	77999 to 78135	77	Fail			Low amplification
	crAss266-Rev	ATGATTATAGTGGAAAGGCTCGTAA						
crAss267	crAss267-For	AGATATTTACGAGCCTTTCCACT	78105 to 78237	77	Fail			Low amplification
	crAss267-Rev	AACAAATGTTTACTGCTTCACAGA						
crAss268	crAss268-For	TTTACGAGCCTTTCCACTATAATCA	78110 to 78223	77	Fail			Low amplification
	crAss268-Rev	GCTTCACAGATTACTAATGCAGC						
crAss269	crAss269-For	TCAGGAGTAATCTGACCTTCAAGTA	78258 to 78399	77	Fail			Primer Dimer
	crAss269-Rev	GTGCAGTTGCTGATGAATTTGG						

Table B3 (continued)

crAss270	crAss270-For	TACCAAATTCATCAGCAACTGCACG	78376 to 78485	77	Pass	Fail		Sensitivity, Primer Dimer
	crAss270-Rev	GAAATGGCATCTTGAAGCTGATTCT						
crAss271	crAss271-For	TTCAGAATCAGCTTCAAGATGCCA	78458 to 78575	77	Fail			No Product
	crAss271-Rev	GGAACAGATTGGAGTAGGTTGTGA						
crAss272	crAss272-For	AAGAGCATCTCTATCTTCCATTGT	78581 to 78727	77	Fail			Low amplification
	crAss272-Rev	TGGCAGCCTATTGCTAAACTTC						
crAss273	crAss273-For	AGTTTAGCAATAGGCTGCCA	78708 to 78845	77	Fail			Primer Dimer
	crAss273-Rev	TGGTCAATGGAATGATAGACCTT						
crAss274	crAss274-For	CCCAAAGTTTACCAGTTTCAGGG	79027 to 79136	78	Pass	Fail		Sensitivity, Primer Dimer
	crAss274-Rev	TAGTGGTTCTATTACGCTTCCTGT						
crAss275	crAss275-For	CCTTCTTCATCAACAGGAAGCG	79101 to 79238	78	Fail			Primer Dimer
	crAss275-Rev	GGCGGCATATCAACTTGGTC						
crAss276	crAss276-For	TGCCGCCATAGCAGATTGAA	79232 to 79343	78	Pass	Fail		Spurious Band
	crAss276-Rev	TCTTATGGCACAATATGGACTTGA						
crAss277	crAss277-For	GCATATCAAGTCCATATTGTGCCAT	79315 to 79426	78	Pass	Fail		Sensitivity, Primer Dimer
	crAss277-Rev	ACTGAACAAGTTGAGATTGACCCT						
crAss278	crAss278-For	TCAACTTGTTCAGTTCCATCAGC	79413 to 79561	78	Fail			Primer Dimer
	crAss278-Rev	AGCACTAGAAGCCTTAGAATTGC						
crAss279	crAss279-For	AGCAATTCTAAGGCTTCTAGTGCTT	79538 to 79669	79	Fail			Low amplification, primer dimer
	crAss279-Rev	ATTGCGTATTGCAAGCATGGC						
crAss280	crAss280-For	AGATAATCTTCATCACCACGATACA	79898 to 79996	79	Pass	Fail		Sensitivity, Primer Dimer
	crAss280-Rev	TCGTGACCAAATAGAATATAACGCT						
crAss281	crAss281-For	GCGTTATATTCTATTTGGTCACGA	79973 to 80066	79	Pass	Fail		False Positive, Primer Dimer
	crAss281-Rev	TGGTTAATGCAGCTTCGGAATG						
crAss282	crAss282-For	TTCCGAAGCTGCATTAACCA	80047 to 80160	80	Fail			Low amplification, primer dimer
	crAss282-Rev	CCTTTAGATGATATCGCCAAACTTC						
crAss283	crAss283-For	GAAGTTTGGCGATATCATCTAAAGG	80136 to 80250	80	Fail			Low amplification, primer dimer
	crAss283-Rev	CAAAATGCCTTTGAACAAGGTCT						
crAss284	crAss284-For	AGCAGCTCCAGTAGGAAATGCTAT	80430 to 80554	81	Fail			Low amplification, primer dimer
	crAss284-Rev	AACCTGATAGTGCAGATGCTGGAT						
crAss285	crAss285-For	CAGGTTGTTTCAAAGGATTAAGCGG	80549 to 80698	81	Fail			Primer Dimer
	crAss285-Rev	CAGCTACTGGTTTTCCTGCATA						

Table B3 (continued)

crAss286	crAss286-For	CTCCATATGCAGGAAAACCAGT	80672 to 80796	81	Fail			Primer Dimer
	crAss286-Rev	TGTTGCTGAGGCTAAAATCACTC						
crAss287	crAss287-For	AGAGTGATTTTAGCCTCAGCA	80773 to 80918	81	Fail			No Product
	crAss287-Rev	AATGAACGTAATCGTTGGACTGC						
crAss288	crAss288-For	GCAGTCCAACGATTACGTTC	80896 to 81013	81	Fail			Primer Dimer
	crAss288-Rev	AACCTGCTACTACTTTCTCTGCTA						
crAss289	crAss289-For	CATCTCCGATAGTAAGATTAGCAGA	80972 to 81087	81	Fail			Primer Dimer
	crAss289-Rev	TGGTCGTGAAGCAAATGGTC						
crAss290	crAss290-For	CTGGACCATTTGCTTCACGAC	81065 to 81214	81	Fail			Low amplification, primer dimer
	crAss290-Rev	GCAAAGTCGCTTATGGAGCAG						
crAss291	crAss291-For	GCTCCATTAGAAGTAGCACCA	81272 to 81381	82	Fail			Primer Dimer
	crAss291-Rev	AGAGAAGAAGATTGGGTTAATTGTG						
crAss292	crAss292-For	AACTTAACTTTAGCAGGTTCTTTGA	81392 to 81537	82	Fail			Primer Dimer
	crAss292-Rev	AATCGTGCTGCGAAAGATGC						
crAss293	crAss293-For	GAGCATCTTTCGCAGCACG	81516 to 81655	82	Fail			Low amplification, primer dimer
	crAss293-Rev	TGATAGCGACGGAATAATGCT						
crAss294	crAss294-For	GCCATTATAACTAACTTGAAAGCCT	81604 to 81720	82	Pass	Fail		Spurious Band
	crAss294-Rev	GGTACTGTTAACGGCGGAGA						
crAss295	crAss295-For	CGTTAACAGTACCCGCTGTG	81708 to 81818	82	Fail			Primer Dimer
	crAss295-Rev	TAGTAGAAAACGTCGGACCTG						
crAss296	crAss296-For	AACAGGTCCGACGTTTTCTACT	81796 to 81916	82	Fail			Primer Dimer
	crAss296-Rev	GGCACAACAGATGGGTATGC						
crAss297	crAss297-For	CCCATCTGTTGTGCCAGTTC	81902 to 82049	84	Pass	Fail		False Positive
	crAss297-Rev	GGTGCTAATGTTAAGGAAGCTGTT						
crAss298	crAss298-For	TCTTAAACCATGTAACAGCTTCCTT	82013 to 82131	84	Pass	Fail		Sensitivity
	crAss298-Rev	AATTGAGGTTCTTATTTCTCGTGG						
crAss299	crAss299-For	TCTCCACGAGAAATAAGAACCTCAA	82105 to 82252	84	Pass	Fail		False Positive, Spurious Band
	crAss299-Rev	TTCCTAACGGTATGGCTATGGAT						
crAss300	crAss300-For	CAGTATCCATAGCCATACCGTT	82226 to 82375	84	Pass	Fail		False Positive, Primer Dimer
	crAss300-Rev	AGCGTCTTGCTAAACATCGTC						
crAss301	crAss301-For	AGCCGAATTAATTTCCTGACGA	82338 to 82437	84	Pass	Pass	Fail	False Positive
	crAss301-Rev	TGCTCTTATTAATTCTGACCCATCT						

Table B3 (continued)

crAss302	crAss302-For	AGCTTCTCTAAGTTCTTTAACTCCT	82566 to 82677	84	Pass	Fail		Sensitivity
	crAss302-Rev	TTTCTTTTGGAATCGTAAACGAGAT						
crAss303	crAss303-For	TCTTCGGCTCTAAAACGAAGATAA	82630 to 82778	84	Pass	Pass	Fail	Sensitivity
	crAss303-Rev	GGTCTTGCTCCTAATAATGAAAACT						
crAss304	crAss304-For	AGTTTTCATTATTAGGAGCAAGACC	82754 to 82863	84	Fail			Primer Dimer
	crAss304-Rev	GGATAAGATTGGTAGTTGTGTTCGT						
crAss305	crAss305-For	AGAACTACGAACACAACTACCA	82833 to 83012	84	Fail			Primer Dimer
	crAss305-Rev	GGAGTAGATAAACAGCCTGCTA						
crAss306	crAss306-For	TGTATTAGCAGGCTGTTTATCTACT	82986 to 83097	84	Fail			Primer Dimer
	crAss306-Rev	AAATGAAGCTAATAAAGCAGGAGAA						
crAss307	crAss307-For	CTACCTGCAAACCAATAGAGCCT	83181 to 83290	86	Fail			Primer Dimer
	crAss307-Rev	AGGTATGACTGATATTCCTGCTTG						
crAss308	crAss308-For	CAATCCCCAGCAAGCAGGAA	83257 to 83402	86	Fail			Primer Dimer
	crAss308-Rev	TGCCTATTACTTCTCATCAAGCTG						
crAss309	crAss309-For	GCAGCTTCAGCATCAGTACC	83438 to 83550	86	Fail			Primer Dimer
	crAss309-Rev	GGTGATGGTCTTTCTTATGGACG						
crAss310	crAss310-For	AAAGTAACGTCCATAAGAAAGACCA	83521 to 83668	86	Fail			Primer Dimer
	crAss310-Rev	GAATGTAGCTCTTATGGGTGGC						
crAss311	crAss311-For	TTGCCACCCATAAGAGCTACA	83645 to 83790	86	Fail			Primer Dimer
	crAss311-Rev	GCTGGTATGTTAGAAATTTGTCGTG						
crAss312	crAss312-For	ACAGTCAGAACTTCACCATAAGT	83729 to 83873	86	Fail			False Positive
	crAss312-Rev	TGTGGAAGTCTGAATATAACCGT						
crAss313	crAss313-For	AACGGTTATATTCAGACTTCCACA	83850 to 83965	86	Fail			Primer Dimer
	crAss313-Rev	CCAAAATGATAATGGTGGTACTTCT						
crAss314	crAss314-For	GTTGAACTGTCGCATCTCTTCG	83905 to 84054	86	Fail			Primer Dimer
	crAss314-Rev	CCTGCTGGAATGACTTCTCAAC						
crAss315	crAss315-For	GTCATTCCAGCAGGACCCAT	84041 to 84152	86	Fail			Primer Dimer
	crAss315-Rev	TTGACCCTCAATGGTTGGCT						
crAss316	crAss316-For	TCATTGCCCAATACATACCTTTAGC	84111 to 84260	86	Fail			False Positive
	crAss316-Rev	CTGCTCCTGATGGTGTTACTCA						
crAss317	crAss317-For	TTGACCAGGTTTAGTAACAGCAG	84322 to 84447	86	Fail			Low amplification, primer dimer
	crAss317-Rev	GAAGGTCAAGGTAGTCGTGGT						

Table B3 (continued)

crAss318	crAss318-For	AGGAAAACGGTCATCATCCATAC	84463 to 84611	86	Pass	Fail			False Positive, Spurious Band, Primer Dimer
	crAss318-Rev	TGAAAACATTACGTGTTATTGGACA							
crAss319	crAss319-For	AGCTCCTTTTGTAGATTTACGTTGA	84695 to 84842	88	Fail				Primer Dimer
	crAss319-Rev	TGCTGAAAGACGCGATGAAG							
crAss320	crAss320-For	ATCGCGTCTTTCAGCAGGAG	84827 to 84975	88	Fail				Primer Dimer
	crAss320-Rev	TTCCCGAAACTGTTATTATTGAACG							
crAss321	crAss321-For	TCGGGAATACGATAACCAGCAAT	84969 to 85095	88	Fail				Primer Dimer
	crAss321-Rev	TGCGTGAAGCTAATGCTAAAGAAG							
crAss322	crAss322-For	TAGCTTCACGCATTTCGTTATCAG	85084 to 85231	88	Fail				Low amplification, primer dimer
	crAss322-Rev	ATTGTTCGTGAGGCGTTTAAGGA							
crAss323	crAss323-For	ACTTATCAACATTACCACGACGAT	85180 to 85300	88	Pass	Fail			Spurious Band
	crAss323-Rev	GGCTTTGGTGAACTTCGAGA							
crAss324	crAss324-For	ACCGCTTCTATCTCGAAGTTCA	85271 to 85399	88	Fail				Primer Dimer
	crAss324-Rev	GACGAATTTGCTCAAGCTGGTGT							
crAss325	crAss325-For	CTTGAGCAAATTCGTCACGTTTTA	85384 to 85497	88	Pass	Fail			False Positive, Spurious Band
	crAss325-Rev	GTGTTTCTGTTACTTCAGAAGATGG							
crAss326	crAss326-For	AGTAACAGAAACACCTACAAGTTCT	85484 to 85632	88	Pass	Fail			False Positive
	crAss326-Rev	ACGGTAATCTTATTGACGATAAAGG							
crAss327	crAss327-For	TTCTTTAAGATAAGCAGCAACTTCA	85562 to 85709	88	Fail				Low amplification, primer dimer
	crAss327-Rev	TGACACTGATGCTGAACATGG							
crAss328	crAss328-For	GTCATTCGCTTTGTCATTAGGCTT	85706 to 85854	88	Pass	Fail			Primer Dimer
	crAss328-Rev	GTAAAACAGGGCAGTTAGATGCTG							
crAss329	crAss329-For	TCACTATTCCCATCTCCATTACCA	85776 to 85897	88	Fail				Primer Dimer
	crAss329-Rev	GGTGATGGAAAACCTACTGACG							
crAss330	crAss330-For	AAGTTGTGCTTCAACAGAAGTC	86381 to 86490	91	Pass	Fail			False Positive, Spurious Band
	crAss330-Rev	GTGATGCTGAAAAGAATCAAGCTG							
crAss331	crAss331-For	AGTCTAGCATTTTCCATACGTTCT	86442 to 86554	91	Fail				False Positive
	crAss331-Rev	CTTGATGGTCAAATTGAAGCTATGA							
crAss332	crAss332-For	TCTTTTCAGCATCACTAAGACCA	86476 to 86615	91	Fail				Low amplification, primer dimer
	crAss332-Rev	ACTTGCTAAGATTGATAGAAAAGCA							
crAss333	crAss333-For	CGTTTTTGCAATCTTGACTTTATCA	86318 to 86450	91	Fail				Primer dimer
	crAss333-Rev	TGCTAGACTTAATCTTGAACGTAGT							

Table B3 (continued)

crAss334	crAss334-For	TTGACGAAGTTGTTCTGTTTGTTGA	86624 to 86738	91	Fail			Low amplification, primer dimer
	crAss334-Rev	TCGTGGAGATAATGTTGCTCA						
crAss335	crAss335-For	GCAACATTATCTCCACGAATAGCA	86721 to 86830	91	Fail			Primer dimer
	crAss335-Rev	TCTGTTAAAGAACGTGAGAAACTTG						
crAss336	crAss336-For	GTTCAAGTTTCTCACGTTCTTTAAC	86803 to 86951	91	Fail			False Positive
	crAss336-Rev	TAAACTTGCTTGGATTGATGGTCT						
crAss337	crAss337-For	AAGACCATCAATCCAAGCAAGTT	86927 to 87049	91	Fail			False Positive
	crAss337-Rev	GACGAAGCAGTTATTCGTGGA						
crAss338	crAss338-For	AACTGCTTCGTCAGTAACTCC	87038 to 87162	91	Pass	Fail		Primer Dimer
	crAss338-Rev	TTACGGAACTTGGGCAACTTATTC						
crAss339	crAss339-For	GAATAAGTTGCCCAAGTTCCGT	87139 to 87267	91	Fail			Spurious Bands
	crAss339-Rev	ATCGTATGGCTGCTGATGGT						
crAss340	crAss340-For	CCATACGATATATAGTTTCAGCAGG	87259 to 87437	91	Fail			Primer dimer
	crAss340-Rev	AGAACTTTTGCCTGGTTTTGGA						
crAss341	crAss341-For	TCTTCCAAAACCAGGCAAAAGT	87413 to 87524	91	Pass	Fail		False Positive
	crAss341-Rev	TGGCTCTCGTGCTACAAGTAT						
crAss342	crAss342-For	ACGCCACACCCATTCAATACT	87551 to 87690	91	Fail			Primer dimer
	crAss342-Rev	GGCATACTGTTTGGAGAGGTGA						
crAss343	crAss343-For	ACCATTATTAGCATCACGAGCC	87704 to 87831	91	Fail			Primer dimer
	crAss343-Rev	GCGATAGAGCACTTTTAAATTGGG						
crAss344	crAss344-For	CCCAATTTAAAAGTGCTCTATCGC	87808 to 87948	91	Fail			Primer dimer
	crAss344-Rev	TTGCTGAACGTCGTATGTTGAC						
crAss345	crAss345-For	TTGTTGTTTAGTCAACATACGACG	87917 to 88008	91	Pass	Fail		False Positive, Spurious Band
	crAss345-Rev	GTGTTAGAGATGCTTTTCCTGTTCC						
crAss346	crAss346-For	GCATCTCTAACACTAACAACACG	87996 to 88133	91	Fail			Primer dimer
	crAss346-Rev	TGTTATTGATGACCTTACTGATGCT						
crAss347	crAss347-For	CTTGTCCTTGTGCACTAATATCAT	88144 to 88267	91	Fail			Primer dimer
	crAss347-Rev	ATGCAGTTTGTTAATGAAGGCAATA						
crAss348	crAss348-For	ATTATTGCCTTCATTAACAAACTGC	88241 to 88350	91	Pass	Fail		Sensitivity, Primer Dimer
	crAss348-Rev	TATTTGCTAGAGATGCTGAACTTGG						
crAss349	crAss349-For	AGCTCCAACAATAAAATCATGTGG	88367 to 88476	91	Fail			Primer dimer
	crAss349-Rev	ATCTAACTCGTTTTCCAGCTACT						

Table B3 (continued)

crAss350	crAss350-For	CATAGTAGCTGGAAAACGAGTTAGA	88451 to 88594	91	Fail			False Positive
	crAss350-Rev	GCTGGTATTGCTTGTAAAGCTGATT						
crAss351	crAss351-For	CAGCTTCTATAACGTAATCACAACA	88591 to 88684	91	Fail			False Positive
	crAss351-Rev	GGTAGACTTGATTTTCCTAATCAGC						
crAss352	crAss352-For	CTTCTGTAACTTTCTTACGACTAGC	88756 to 88866	92	Pass	Fail		Sensitivity, Primer Dimer
	crAss352-Rev	CCGAAGGTAACTTTGACCGTA						
crAss353	crAss353-For	GTCAAAGTTACCTTCGGCATTAAAC	88850 to 88999	92	Fail			False Positive
	crAss353-Rev	AAACTAGATGGTCTTCGACTTCT						
crAss354	crAss354-For	GAACTCTTTAAGAAGTCGAAGACCA	88967 to 89079	92	Fail			Low amplification, primer dimer
	crAss354-Rev	AACCTTTATATGTTTGGGATTCTGC						
crAss355	crAss355-For	TCTTTAACAGCAGAATCCCAAACA	89046 to 89156	92	Pass	Fail		Spurious Band, Primer Dimer
	crAss355-Rev	TGGACTTGTAGAAATAAATCGTGGT						
crAss356	crAss356-For	CACCACGATTTATTTCTACAAGTCC	89131 to 89243	92	Fail			Primer Dimer
	crAss356-Rev	TGCTGCATATTATGGACGTACTGA						
crAss357	crAss357-For	TGCAGCACACAACTTAGGTTTA	89237 to 89381	92	Fail			Primer Dimer
	crAss357-Rev	TCCCGGTACTTATGTTGCAGT						
crAss358	crAss358-For	TGCAACATAAGTACCGGGAAGA	89363 to 89479	92	Pass	Fail		Primer Dimer
	crAss358-Rev	AGACGTGGTAACGAAGACCC						
crAss359	crAss359-For	AGTTCCATCTTCAAATAGTTCACCA	89564 to 89711	92	Pass	Fail		Sensitivity
	crAss359-Rev	TCAGCGTGCTTTATTTCCTGC						
crAss360	crAss360-For	AGCACTACTAAAAGATTCAGCAGGA	89672 to 89784	92	Fail			False Positive
	crAss360-Rev	AGCTTGCTGCACGAGAAAGA						
crAss361	crAss361-For	TCAGCAGGAAATAAAGCACGC	89688 to 89797	92	Fail			Primer Dimer
	crAss361-Rev	CGAATAGGACTTAAGCTTGCTGC						
crAss362	crAss362-For	TCGTGCAGCAAGCTTAAGTCC	89771 to 89900	92	Fail			No Product
	crAss362-Rev	TGCTCGTAATGAAGTTTGTGGATTC						
crAss363	crAss363-For	ATCCACAAACTTCATTACGAGCA	89878 to 90005	92	Fail			No Product
	crAss363-Rev	GGCTTGGGGAACTGCTACTG						
crAss364	crAss364-For	AGCAGTTCCCCAAGCCATAA	89990 to 90137	92	Fail			False Positive
	crAss364-Rev	TGACTGTGCTATTGGTAAAGATGC						
crAss365	crAss365-For	TGCATTGTAGATAGCTCTTCAACTT	90078 to 90199	92	Fail			Low amplification, primer dimer
	crAss365-Rev	GGAGTTGAAGCTGATGATAGTTGG						

Table B3 (continued)

crAss366	crAss366-For	GACCAACTATCATCAGCTTCAACTC	90174 to 90284	92	Fail			Primer Dimer
	crAss366-Rev	GTTCTTTGAAGAAAAGACTCCGTTT						
crAss367	crAss367-For	TCTTTTAAACGGAGTCTTTTCTTCA	90254 to 90363	92	Fail			No Product
	crAss367-Rev	TTGTTATTCATGTTGCGGCTG						
crAss368	crAss368-For	TCATTAGAACTATCAGCAGCCAT	90387 to 90508	92	Fail			No Product
	crAss368-Rev	GATGCTCAATTTTGGACTTGGC						
crAss369	crAss369-For	ACCTCCACGTCGAGTTTTATCA	90422 to 90559	92	Fail			Primer Dimer
	crAss369-Rev	ACTGAACATGGAGCTACTGC						
crAss370	crAss370-For	GCAGTAGCTCCATGTTCAGTAAC	90540 to 90679	92	Pass	Fail		False Positive
	crAss370-Rev	TCTGCTCCTTGTTGGCAAAATC						
crAss371	crAss371-For	GGATTTTGCCAACAAGGAGCA	90657 to 90788	92	Fail			Primer Dimer
	crAss371-Rev	AGCTGCTAGAACATATCAAGCCA						
crAss372	crAss372-For	TGGCTTGATATGTTCTAGCAGCTT	90766 to 90906	92	Fail			Primer Dimer
	crAss372-Rev	AGAAAGCTAGTGATTGTGGATGGT						
crAss373	crAss373-For	GGGTCATACCATCCACAATCACT	90876 to 90970	92	Pass	Fail		False Positive
	crAss373-Rev	ATGGCTGATGGTAAATATCCTTTTC						
crAss374	crAss374-For	AGTTGTATCAGTAGTGTTGACTCCC	91265 to 91359	94	Fail			No Product
	crAss374-Rev	CAGTGCTACTTGGGAGTTATTCAAG						
crAss375	crAss375-For	AAGCAAATCAAGATTCCATCTACC	91642 to 91770	94	Pass	Pass	Fail	False Positive
	crAss375-Rev	TTTAATAGTCAGAGAGTTGCTGAAC						
crAss376	crAss376-For	TCATCAAGTTCAAATACATCCCAAT	91772 to 91921	94	Fail			Primer Dimer
	crAss376-Rev	GCAAGGTTGGACTATGATTGATTCT						
crAss377	crAss377-For	AATCAATCATAGTCCAACCTTGCC	91899 to 91988	94	Fail			No Product
	crAss377-Rev	TAAACGATAAAGATGATGCTGGTGG						
crAss378	crAss378-For	GTTTCCTTTTCCGGCAATATCA	92296 to 92416	95	Fail			Primer Dimer
	crAss378-Rev	TAGAACTCAAGATTCTCTTAACGCT						
crAss379	crAss379-For	TCAACAGCAAAAGCATACATACTAA	92842 to 92973	96	Fail			False Positive
	crAss379-Rev	AGTTAAACCTGCACAACAAACTAAT						
crAss380	crAss380-For	CAAACATTAGTTTGTTGTGCAGGT	92944 to 93053	96	Fail			No Product
	crAss380-Rev	AATTGTGCTGTTGTAGGCGAT						
crAss381	crAss381-For	TCCAATCGCCTACAACAGCA	93029 to 93171	96	Fail			Primer Dimer
	crAss381-Rev	TGATGCGTTTAAAATGTACACTGA						

Table B3 (continued)

crAss382	crAss382-For	AGTAACCATTACACCATCTTCTCG	93264 to 93386	96	Pass	Fail		Sensitivity, Primer Dimer
	crAss382-Rev	AAAGAAGCTTGTGAACGTGGAA						
crAss383	crAss383-For	GGTCAGCATCCAATCCAATTCC	94266 to 94378	98	Pass	Fail		Primer Dimer
	crAss383-Rev	TGCTGCTTATGCTAATCTTTGGC						
crAss384	crAss384-For	CGATACCAATTCTTTTGAATAGTGC	94577 to 94669	98	Pass	Fail		False Positive, Spurious Band
	crAss384-Rev	TGAAAGTCTTGAAAAAGAAGCTAGT						

Table B4: Summary of round three candidate end-point primer set testing[a]

Primer Set	Non-Human Host Distribution		Sewage Distribution	LOD_{95}	
	Specificity	*False Positives*		*0.01ng*	*0.001ng*
crAss028	100%	None	92.6%	---	---
crAss056	100%	None	100%	100%	45%
crAss064	100%	None	100%	100%	52.5%
crAss301	96%	Pig (n=5); Goose (n=2); Dog (n=1)	---	---	---
crAss303	100%	None	92.6%	---	---
crAss375	95%	Pig (n=3); Horse (n=3); Goose (n=1); Deer (n=1); Dog (n=1); Cow (n=1)	---	---	---

[a] Specificity, calculated as the percent of negative non-target animal samples; Sewage Distribution, calculated as the percent of positive sewage samples; LOD_{95}, defined as the lowest dilution concentration where a minimum of 95% (38 of 40) of reactions yielded an amplification product of the expected size; a "---" indicates that primer set was not evaluated.

Table B5: Summary of crAssphage and bacterial human-associated qPCR calibration model performance metrics[a]

Assay	Master Slope	Y-Intercept Range	*E*	LLOQ Range	CV
CPQ_056	-3.47±0.03	37.7 to 39.3	0.94	0.82 to 0.94	1.88%
CPQ_064	-3.39±0.03	39.4 to 40.7	0.97	0.86 to 0.97	0.81%
HF183/BacR287	-3.48±0.03	35.9 to 36.9	0.94	0.86 to 0.95	1.62%
HumM2	-3.36±0.05	37.1 to 38.2	0.99	0.85 to 0.96	1.79%

[a]Master Slope, average and standard deviation of six calibration curves ran for each assay; Y-Intercept Range, range of the six y-intercept measurements run on each individual plate; E, amplification efficiency, calculated as $E=10^{(-1/slope)}-1$; LLOQ Range, Lower Limit of Quantification Range, expressed in $\log_{10}$ copies/reaction, defined as the upper bound of the 95% credible interval from repeated measures of the 10 copy per reaction standard curve dilutions; CV, % coefficient of variation, calculated from standard data at 10 copies per reaction.

A.

```
crAss_056            1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Sewage_056_A_(13)    1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Sewage_056_B_(1)     1 AGTGCAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Sewage_056_C_(1)     1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACAATTTACGTGATGTAAC
Water_056_A_(24)     1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Water_056_D_(2)      1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Water_056_E_(2)      1 AGTACAAACTCCTAA-AAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC
Water_056_F_(1)      1 AGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAAC

crAss_056           61 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Sewage_056_A_(13)   61 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Sewage_056_B_(1)    61 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Sewage_056_C_(1)    61 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Water_056_A_(24)    61 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Water_056_D_(2)     61 TCGTAAAAAGTTTGATGAACGTACTGATTGCAATAAAGCTAA
Water_056_E_(2)     60 TCGTAAAAAGTTTGATGAACGTACTGATTGTAATAAAGCTAA
Water_056_F_(1)     61 TCGTAAAAAGTTTGATGAACGTACTAATTGTAATAAAGCTAA
```

B.

```
crAss_064            1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Sewage_064_A_(13)    1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Sewage_064_B_(1)     1 CTGAGATTGTTCATAAGCAAATTGATATTTCTATTAAGAGTCAATTTCTATTTGTTCTTA
Sewage_064_C_(1)     1 CTGAAATTGTTCATAAGCAAATTGAGATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_A_(22)     1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_D_(6)      1 CTGAGATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_E_(1)      1 CTGAAATTGTTCATAAGCAAATTGATATTTCTGTTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_F_(1)      1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_G_(1)      1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA
Water_064_H_(1)      1 CTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTA

crAss_064           61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Sewage_064_A_(13)   61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Sewage_064_B_(1)    61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Sewage_064_C_(1)    61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_A_(22)    61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_D_(6)     61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_E_(1)     61 AACATATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_F_(1)     61 AACGTATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_G_(1)     61 AGCATATTCGTTATACTTTTAGAAATATTATTTAT
Water_064_H_(1)     61 AACATATTCGTTATACTTTTAGAAGTATTATTTAT
```

Figure B1: CrAssphage sequence verification alignment with sequences isolated from primary influent sewage and a human fecal pollution impact environmental water sample to the crAss056 amplicon (A) and the crAss064 amplicon (B). The letters 'A-H' represent the variant designation and the number in parenthesis is the sequence count for respective sample source/variant combination. Differences are highlighted in black.

APPENDIX C

CHAPTER 5 – SUPPLEMENTAL INFORMATION

Table C1 Spearman's rank correlation coefficients matrix for culturable indicators, qPCR indicators, and chemical parameters.

	E. coli	***Enterococci***	**Somatic coliphage**	**CPQ_056**	**CPQ_064**	**HF183**	**HPyV**	**TOC**	**pH**	**TDS**	**Turbidity**
E. coli		0.934	0.581	0.577	0.734	0.876	0.722	0.839	-0.481	-0.598	0.763
Enterococci	0.934		0.466	0.420	0.656	0.862	0.670	0.826	-0.443	-0.585	0.795
Somatic coliphage	0.581	0.466		0.669	0.541	0.539	0.378	0.469	-0.291	-0.570	0.506
CPQ_056	0.577	0.420	0.669		0.675	0.420	0.275	0.348	-0.395	-0.378	0.334
CPQ_064	0.734	0.656	0.541	0.675		0.784	0.716	0.675	-0.544	-0.524	0.430
HF183	0.876	0.862	0.539	0.420	0.784		0.774	0.777	-0.498	-0.666	0.732
HPyV	0.722	0.670	0.378	0.275	0.716	0.774		0.672	-0.256	-0.514	0.483
TOC	0.839	0.826	0.469	0.348	0.675	0.777	0.672		-0.397	-0.740	0.714
pH	-0.481	-0.443	-0.291	-0.395	-0.544	-0.498	-0.256	-0.397		0.425	-0.369
TDS	-0.598	-0.585	-0.570	-0.378	-0.524	-0.666	-0.514	-0.740	0.425		-0.708
Turbidity	0.762	0.795	0.506	0.334	0.430	0.732	0.483	0.714	-0.369	-0.708	

Table C2 P-values for Spearman's rank correlation coefficients matrix for culturable indicators, qPCR indicators, and chemical parameters.

	E. coli	***Enterococci***	**Somatic coliphage**	**CPQ_056**	**CPQ_064**	**HF183**	**HPyV**	**TOC**	**pH**	**TDS**	**Turbidity**
E. coli		1.60 E -14	6.16 E -4	6.86 E - 4	2.63 E -6	1.11 E -10	6.57 E -06	3.83 E -09	6.11 E -3	3.77 E -4	6.18 E -07
Enterococci	1.60 E -14		8.22 E -3	0.0186	6.06 E -5	4.63 E -10	5.15 E -5	1.04 E -08	0.0126	5.41 E -4	9.17 E -08
Somatic coliphage	6.16 E -4	8.22 E -3		3.94 E -5	1.69 E -3	1.75 E -3	0.0397	7.73 E -3	0.113	8.26 E -4	3.72 E -3
CPQ_056	6.86 E - 4	0.0186	3.94 E -5		3.11 E - 5	0.0187	0.141	0.0554	0.0279	0.036	0.0667
CPQ_064	2.63 E -6	6.06 E -5	1.69 E -3	3.11 E - 5		1.83 E -07	8.59 E -06	3.16 E -7	1.58 E - 3	2.49 E -3	0.0158
HF183	1.11 E -10	4.63 E -10	1.75 E -3	0.0187	1.83 E -07		5.12 E -07	2.81 E -07	4.33 E -3	4.37 E -5	2.93 E -06
HPyV	6.57 E -06	5.15 E -5	0.0397	0.141	8.59 E -06	5.12 E -07		4.83 E -5	0.171	3.63 E -3	6.90 E -3
TOC	3.83 E -09	1.04 E -08	7.73 E -3	0.0554	3.16 E -7	2.81 E -07	4.83 E -5		0.0269	1.98 E -06	6.47 E -06
pH	6.11 E -3	0.0126	0.113	0.0279	1.58 E - 3	4.33 E -3	0.171	0.0269		0.017	0.0412
TDS	3.77 E -4	5.41 E -4	8.26 E -4	0.036	2.49 E -3	4.37 E -5	3.63 E -3	1.98 E -06	0.017		8.31 E -06
Turbidity	6.18 E -07	9.17 E -08	3.72 E -3	0.0667	0.0158	2.93 E -06	6.90 E -3	6.47 E -06	0.0412	8.31 E -06	

BIBLIOGRAPHY

1. Sinclair, R., E. Jones, and C.P. Gerba, *Viruses in recreational water-borne disease outbreaks: A review.* Journal of Applied Microbiology, 2009. **107**(6): p. 1769-1780.

2. Soller, J.A., et al., *Estimating the primary etiologic agents in recreational freshwaters impacted by human sources of faecal contamination.* Water Research, 2010. **44**(16): p. 4736-4747.

3. Bartsch, S.M., et al., *Global Economic Burden of Norovirus Gastroenteritis.* PloS one, 2016. **11**(4): p. e0151219.

4. Hall, A.J., et al., *Norovirus disease in the United States.* Emerg Infect Dis, 2013. **19**(8): p. 1198-205.

5. Lopman, B., et al., *Environmental transmission of norovirus gastroenteritis.* Current opinion in virology, 2012. **2**(1): p. 96-102.

6. Lodder, W. and A. de Roda Husman, *Presence of noroviruses and other enteric viruses in sewage and surface waters in The Netherlands.* Applied and Environmental microbiology, 2005. **71**(3): p. 1453-1461.

7. Fankhauser, R.L., et al., *Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United States.* The Journal of infectious diseases, 1998. **178**(6): p. 1571-1578.

8. Rodriguez-Lazaro, D., et al., *Virus hazards from food, water and other contaminated environments.* FEMS microbiology reviews, 2012. **36**(4): p. 786-814.

9. Harwood, V.J., et al., *Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes.* FEMS microbiology reviews, 2014. **38**(1): p. 1-40.

10. Green, H.C., et al., *Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water.* Applied and environmental microbiology, 2012. **78**(2): p. 503-510.

11. Green, H.C., et al., *Development of Rapid Canine Fecal Source Identification PCR-Based Assays.* Environmental science & technology, 2014. **48**(19): p. 11453-11461.

12. Shanks, O.C., et al., *Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples.* Applied and environmental microbiology, 2006. **72**(6): p. 4054-4060.

13. Mieszkin, S., et al., *Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers.* Applied and environmental microbiology, 2009. **75**(10): p. 3045-3054.

14. Lu, J., J. Santo Domingo, and O.C. Shanks, *Identification of chicken-specific fecal microbial sequences using a metagenomic approach.* Water research, 2007. **41**(16): p. 3561-3574.

15. Shanks, O.C., et al., *Identification of bacterial DNA markers for the detection of human fecal pollution in water.* Applied and environmental microbiology, 2007. **73**(8): p. 2416-2422.

16. Shanks, O.C., et al., *Quantitative PCR for genetic markers of human fecal pollution.* Applied and Environmental Microbiology, 2009. **75**(17): p. 5507-5513.

17. Green, H.C., et al., *Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples.* Applied and environmental microbiology, 2014. **80**(10): p. 3086-3094.

18. Soller, J.A., et al., *Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination.* Water Research, 2010. **44**(16): p. 4674-4691.

19. Boehm, A. and J. Soller, *Risks associated with recreational waters: pathogens and fecal indicators.* Encyclopedia of Sustainability Science and Technology, 2011.

20. Heim, A., et al., *Rapid and quantitative detection of human adenovirus DNA by real-time PCR.* Journal of medical virology, 2003. **70**(2): p. 228-239.

21. Jothikumar, N., et al., *Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41.* Applied and environmental microbiology, 2005. **71**(6): p. 3131-3136.

22. Jothikumar, N., et al., *Rapid and sensitive detection of noroviruses by using TaqMan-based one-step reverse transcription-PCR assays and application to naturally contaminated shellfish samples.* Applied and environmental microbiology, 2005. **71**(4): p. 1870-1875.

23. McQuaig, S.M., et al., *Quantification of human polyomaviruses JC virus and BK virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples.* Applied and environmental microbiology, 2009. **75**(11): p. 3379-3388.

24. Rosario, K., et al., *Pepper mild mottle virus as an indicator of fecal pollution.* Applied and environmental microbiology, 2009. **75**(22): p. 7261-7267.

25. Duizer, E., et al., *Laboratory efforts to cultivate noroviruses.* Journal of General Virology, 2004. **85**(1): p. 79-87.

26. Ikner, L.A., C.P. Gerba, and K.R. Bright, *Concentration and recovery of viruses from water: a comprehensive review.* Food and environmental virology, 2012. **4**(2): p. 41-67.

27. Cashdollar, J. and L. Wymer, *Methods for primary concentration of viruses from water samples: a review and meta‐analysis of recent studies.* Journal of applied microbiology, 2013. **115**(1): p. 1-11.

28. Ahmed, W., et al., *Comparison of concentration methods for quantitative detection of sewage-associated viral markers in environmental waters.* Applied and environmental microbiology, 2015. **81**(6): p. 2042-2049.

29. Jofre, J., et al., *Evaluation of< i> Bacteroides fragilis</i> Bacteriophages as Indicators of the Virological Quality of Water.* Water Science & Technology, 1986. **18**(10): p. 167-173.

30. Leclerc, H., et al., *Bacteriophages as indicators of enteric viruses and public health risk in groundwaters.* Journal of applied microbiology, 2000. **88**(1): p. 5-21.

31. Weinbauer, M.G., *Ecology of prokaryotic viruses.* FEMS microbiology reviews, 2004. **28**(2): p. 127-181.

32. Dutilh, B.E., et al., *A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes.* Nature communications, 2014. **5**.

33. Wade, T.J., et al., *Do US Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis.* Environmental Health Perspectives, 2003. **111**(8): p. 1102.

34. Brown, K., et al., *Estimating probability of illness due to swimming in recreational water with a mixture of human-and gull-associated microbial source tracking markers.* Environmental Science: Processes & Impacts, 2017.

35. Cabelli, V.J., et al., *A marine recreational water quality criterion consistent with indicator concepts and risk analysis.* Journal (Water Pollution Control Federation), 1983: p. 1306-1314.

36. Ishii, S., et al., *Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds.* Applied and environmental microbiology, 2006. **72**(1): p. 612-621.

37. Badgley, B.D., F.I. Thomas, and V.J. Harwood, *Quantifying environmental reservoirs of fecal indicator bacteria associated with sediment and submerged aquatic vegetation.* Environmental microbiology, 2011. **13**(4): p. 932-942.

38. Whitman, R.L., et al., *Occurrence of Escherichia coli and enterococci in Cladophora (Chlorophyta) in nearshore water and beach sand of Lake Michigan.* Applied and Environmental Microbiology, 2003. **69**(8): p. 4714-4719.

39. Litton, R.M., et al., *Evaluation of chemical, molecular, and traditional markers of fecal contamination in an effluent dominated urban stream.* Environmental science & technology, 2010. **44**(19): p. 7369-7375.

40. Harwood, V.J., et al., *Isolation of fecal coliform bacteria from the diamondback terrapin (Malaclemys terrapin centrata).* Applied and environmental microbiology, 1999. **65**(2): p. 865-867.

41. Souza, V., et al., *Genetic Structure of Natural Populations ofEscherichia coli in Wild Hosts on Different Continents.* Applied and Environmental Microbiology, 1999. **65**(8): p. 3373-3385.

42. Harwood, V.J., et al., *Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection.* Applied and Environmental Microbiology, 2005. **71**(6): p. 3163-3170.

43. Bradshaw, J.K., et al., *Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed.* Water research, 2016. **101**: p. 498-509.

44. Boehm, A.B., et al., *Covariation and photoinactivation of traditional and novel indicator organisms and human viruses at a sewage-impacted marine beach.* Environmental science & technology, 2009. **43**(21): p. 8046-8052.

45. Korajkic, A., M. Brownell, and V. Harwood, *Investigation of human sewage pollution and pathogen analysis at Florida Gulf coast beaches.* Journal of applied microbiology, 2011. **110**(1): p. 174-183.

46. Wanjugi, P., et al., *Differential decomposition of bacterial and viral fecal indicators in common human pollution types.* Water research, 2016. **105**: p. 591-601.

47. Korajkic, A., et al., *Biotic interactions and sunlight affect persistence of fecal indicator bacteria and microbial source tracking genetic markers in the upper Mississippi river.* Applied and environmental microbiology, 2014. **80**(13): p. 3952-3961.

48. Foy, H.M., M.K. Cooney, and J.B. Hatlen, *Adenovirus type 3 epidemic associated with intermittent chlorination of a swimming pool.* Archives of Environmental Health: An International Journal, 1968. **17**(5): p. 795-802.

49. Papapetropoulou, M. and A. Vantarakis, *Detection of adenovirus outbreak at a municipal swimming pool by nested PCR amplification.* Journal of Infection, 1998. **36**(1): p. 101-103.

50. Hauri, A., et al., *An outbreak of viral meningitis associated with a public swimming pond.* Epidemiology & Infection, 2005. **133**(2): p. 291-298.

51. Rose, J.B., et al., *Occurrence of rotaviruses and enteroviruses in recreational waters of Oak Creek, Arizona.* Water research, 1987. **21**(11): p. 1375-1381.

52. Soller, J.A., et al., *Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody.* Water research, 2014. **66**: p. 254-264.

53. Wade, T.J., et al., *Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness.* Environmental health perspectives, 2006. **114**(1): p. 24.

54. Stoeckel, D.M. and V.J. Harwood, *Performance, design, and analysis in microbial source tracking studies.* Applied and Environmental Microbiology, 2007. **73**(8): p. 2405-2415.

55. Boehm, A.B., et al., *Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study.* Water research, 2013. **47**(18): p. 6812-6828.

56. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* nature, 2011. **473**(7346): p. 174-180.

57. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* science, 2005. **308**(5728): p. 1635-1638.

58. Bernhard, A.E. and K.G. Field, *A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA.* Applied and environmental microbiology, 2000. **66**(10): p. 4571-4574.

59. Seurinck, S., et al., *Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater.* Environmental Microbiology, 2005. **7**(2): p. 249-259.

60. Haugland, R.A., et al., *Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected< i> Bacteroidales</i> species and human fecal waste by qPCR.* Systematic and applied microbiology, 2010. **33**(6): p. 348-357.

61. Shanks, O.C., et al., *Performance of PCR-based assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples.* Environmental science & technology, 2010. **44**(16): p. 6281-6288.

62. Jenkins, M.W., et al., *Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays.* Water Research, 2009. **43**(19): p. 4956-4966.

63. Ahmed, W., et al., *Evaluation of multiple sewage-associated Bacteroides PCR markers for sewage pollution tracking.* water research, 2009. **43**(19): p. 4872-4877.

64. Ebentier, D.L., et al., *Evaluation of the repeatability and reproducibility of a suite of qPCR-based microbial source tracking methods.* Water research, 2013. **47**(18): p. 6839-6848.

65. Kildare, B.J., et al., *16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach.* Water research, 2007. **41**(16): p. 3701-3715.

66. Reischer, G.H., et al., *A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area.* Letters in applied microbiology, 2007. **44**(4): p. 351-356.

67. Reischer, G.H., et al., *Performance characteristics of qPCR assays targeting human-and ruminant-associated Bacteroidetes for microbial source tracking across sixteen countries on six continents.* Environmental science & technology, 2013. **47**(15): p. 8548-8556.

68. Layton, A., et al., *Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water.* Applied and environmental microbiology, 2006. **72**(6): p. 4214-4224.

69. Silkie, S.S. and K.L. Nelson, *Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces.* Water research, 2009. **43**(19): p. 4860-4871.

70. Okabe, S., et al., *Quantification of host-specific Bacteroides–Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater.* Applied Microbiology and Biotechnology, 2007. **74**(4): p. 890-901.

71. Lee, C.S. and J. Lee, *Evaluation of new gyrB-based real-time PCR system for the detection of B. fragilis as an indicator of human-specific fecal contamination.* Journal of microbiological methods, 2010. **82**(3): p. 311-318.

72. Lee, D.-Y., et al., *Quantitative identification of fecal water pollution sources by TaqMan real-time PCR assays using Bacteroidales 16S rRNA genetic markers.* Applied microbiology and biotechnology, 2010. **88**(6): p. 1373-1383.

73. Xu, J., et al., *A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.* Science, 2003. **299**(5615): p. 2074-2076.

74. Yampara-Iquise, H., et al., *Use of a Bacteroides thetaiotaomicron‐specific α‐1‐6, mannanase quantitative PCR to detect human faecal pollution in water.* Journal of applied microbiology, 2008. **105**(5): p. 1686-1693.

75. Gourmelon, M., et al., *Development of microbial and chemical MST tools to identify the origin of the faecal pollution in bathing and shellfish harvesting waters in France.* Water research, 2010. **44**(16): p. 4812-4824.

76. Gómez-Doñate, M., et al., *New molecular quantitative PCR assay for detection of host-specific Bifidobacteriaceae suitable for microbial source tracking.* Applied and environmental microbiology, 2012. **78**(16): p. 5788-5795.

77. Yahya, M., et al., *Comparison of the Performance of Different Microbial Source Tracking Markers among European and North African Regions.* Journal of Environmental Quality, 2017.

78. Ahmed, W., et al., *A real‐time polymerase chain reaction assay for quantitative detection of the human‐specific enterococci surface protein marker in sewage and environmental waters.* Environmental microbiology, 2008. **10**(12): p. 3255-3264.

79. Scott, T.M., et al., *Potential use of a host associated molecular marker in Enterococcus faecium as an index of human fecal pollution.* Environmental science & technology, 2005. **39**(1): p. 283-287.

80. Whitman, R.L., et al., *Incidence of the enterococcal surface protein (esp) gene in human and animal fecal sources.* Environmental science & technology, 2007. **41**(17): p. 6090-6095.

81. Boehm, A.B., J.A. Soller, and O.C. Shanks, *Human-associated fecal quantitative polymerase chain reaction measurements and simulated risk of gastrointestinal illness in recreational waters contaminated with raw sewage.* Environmental Science & Technology Letters, 2015. **2**(10): p. 270-275.

82. Ravaliya, K., et al., *Use of bacteroidales microbial source tracking to monitor fecal contamination in fresh produce production.* Applied and environmental microbiology, 2014. **80**(2): p. 612-617.

83. Sidhu, J., et al., *Sewage pollution in urban stormwater runoff as evident from the widespread presence of multiple microbial and chemical source tracking markers.* Science of the Total Environment, 2013. **463**: p. 488-496.

84. Gordon, K.V., et al., *Relationship of human-associated microbial source tracking markers with Enterococci in Gulf of Mexico waters.* Water research, 2013. **47**(3): p. 996-1004.

85. Srinivasan, S., et al., *Escherichia coli, enterococci, and Bacteroides thetaiotaomicron qPCR signals through wastewater and septage treatment.* water research, 2011. **45**(8): p. 2561-2572.

86. Schmitz, B., et al., *Virus reduction during advanced Bardenpho and conventional wastewater treatment processes.* Environmental science & technology, 2016. **50**.

87. Okoh, A.I., T. Sibanda, and S.S. Gusha, *Inadequately treated wastewater as a source of human enteric viruses in the environment.* International journal of environmental research and public health, 2010. **7**(6): p. 2620-2637.

88. Given, S., L.H. Pendleton, and A.B. Boehm, *Regional public health cost estimates of contaminated coastal waters: a case study of gastroenteritis at Southern California beaches*. 2006, ACS Publications.

89. Shuval, H., *Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment.* Journal of Water and Health, 2003. **1**(2): p. 53-64.

90. Bibby, K., *Metagenomic identification of viral pathogens.* Trends in biotechnology, 2013. **31**(5): p. 275-279.

91. Wolf, S., J. Hewitt, and G.E. Greening, *Viral multiplex quantitative PCR assays for tracking sources of fecal contamination.* Applied and environmental microbiology, 2010. **76**(5): p. 1388-1394.

92. da Silva, A.K., et al., *Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: different behaviors of genogroups I and II.* Applied and Environmental Microbiology, 2007. **73**(24): p. 7891-7897.

93. Kageyama, T., et al., *Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR.* Journal of clinical microbiology, 2003. **41**(4): p. 1548-1557.

94. Harwood, V.J., et al., *Performance of viruses and bacteriophages for fecal source determination in a multi-laboratory, comparative study.* Water research, 2013. **47**(18): p. 6929-6943.

95. Hewitt, J., et al., *Evaluation of human adenovirus and human polyomavirus as indicators of human sewage contamination in the aquatic environment.* Water research, 2013. **47**(17): p. 6750-6761.

96. Nordgren, J., et al., *Prevalence of norovirus and factors influencing virus concentrations during one year in a full-scale wastewater treatment plant.* Water research, 2009. **43**(4): p. 1117-1125.

97. Xu, W., M.C. McDonough, and D.D. Erdman, *Species-specific identification of human adenoviruses by a multiplex PCR assay.* Journal of Clinical Microbiology, 2000. **38**(11): p. 4114-4120.

98. Pina, S., et al., *Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses.* Applied and Environmental Microbiology, 1998. **64**(9): p. 3376-3382.

99. Donaldson, K., D.W. Griffin, and J. Paul, *Detection, quantitation and identification of enteroviruses from surface waters and sponge tissue from the Florida Keys using real-time RT–PCR.* Water research, 2002. **36**(10): p. 2505-2514.

100. Fuhrman, J.A., X. Liang, and R.T. Noble, *Rapid detection of enteroviruses in small volumes of natural waters by real-time quantitative reverse transcriptase PCR.* Applied and environmental microbiology, 2005. **71**(8): p. 4523-4530.

101. Monpoeho, S., et al., *Best viral elution method available for quantification of enteroviruses in sludge by both cell culture and reverse transcription-PCR.* Applied and environmental microbiology, 2001. **67**(6): p. 2484-2488.

102. Walters, S.P., K.M. Yamahara, and A.B. Boehm, *Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: implications for their use in assessing risk in recreational waters.* Water Research, 2009. **43**(19): p. 4929-4939.

103. Ng, T.F.F., et al., *High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage.* Journal of virology, 2012. **86**(22): p. 12161-12175.

104. Polo, C., et al., *Prevalence and patterns of polyomavirus urinary excretion in immunocompetent adults and children.* Clinical Microbiology and Infection, 2004. **10**(7): p. 640-644.

105. Gardner, S., et al., *New human papovavirus (BK) isolated from urine after renal transplantation.* The Lancet, 1971. **297**(7712): p. 1253-1257.

106. McQuaig, S.M., et al., *Detection of human-derived fecal pollution in environmental waters by use of a PCR-based human polyomavirus assay.* Applied and environmental microbiology, 2006. **72**(12): p. 7567-7574.

107. Staley, C., et al., *Performance of two quantitative PCR methods for microbial source tracking of human sewage and implications for microbial risk assessment in recreational waters.* Applied and environmental microbiology, 2012. **78**(20): p. 7317-7326.

108. Harwood, V.J., et al., *Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico.* water research, 2009. **43**(19): p. 4812-4819.

109. Symonds, E., et al., *Microbial source tracking in shellfish harvesting waters in the Gulf of Nicoya, Costa Rica.* Water research, 2017. **111**: p. 177-184.

110. McQuaig, S., J. Griffith, and V.J. Harwood, *Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution.* Applied and environmental microbiology, 2012. **78**(18): p. 6423-6432.

111. Zhang, T., et al., *RNA viral community in human feces: prevalence of plant pathogenic viruses.* PLoS biology, 2005. **4**(1): p. e3.

112. Wetter, C., et al., *Pepper mild mottle virus, a tobamovirus infecting pepper cultivars in Sicily.* Phytopathology, 1984. **74**(4): p. 405-410.

113. Alonso, E., et al., *Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistance-breaking tobamovirus in pepper.* The Journal of general virology, 1991. **72**: p. 2875-2884.

114. Hamza, I.A., et al., *Evaluation of pepper mild mottle virus, human picobirnavirus and Torque teno virus as indicators of fecal contamination in river water.* Water research, 2011. **45**(3): p. 1358-1368.

115. Griffith, J.F., S.B. Weisberg, and C.D. McGee, *Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples.* Journal of Water and Health, 2003. **1**(4): p. 141-151.

116. Ahmed, W., et al., *Distributions of fecal markers in wastewater from different climatic zones for human fecal pollution tracking in Australian surface waters.* Applied and environmental microbiology, 2016. **82**(4): p. 1316-1323.

117. Ahmed, W., et al., *Comparison of molecular markers to detect fresh sewage in environmental waters.* Water research, 2009. **43**(19): p. 4908-4917.

118. EPA, *Review of coliphages as possible indicators of fecal contamination for ambient water quality.* 2015.

119. Hsu, F.-C., et al., *Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes.* Applied and environmental microbiology, 1995. **61**(11): p. 3960-3966.

120. Kirs, M. and D.C. Smith, *Multiplex quantitative real-time reverse transcriptase PCR for F+-specific RNA coliphages: a method for use in microbial source tracking.* Applied and environmental microbiology, 2007. **73**(3): p. 808-814.

121. Wolf, S., et al., *Detection and characterization of F+ RNA bacteriophages in water and shellfish: application of a multiplex real-time reverse transcription PCR.* Journal of virological methods, 2008. **149**(1): p. 123-128.

122. Ogorzaly, L. and C. Gantzer, *Development of real-time RT-PCR methods for specific detection of F-specific RNA bacteriophage genogroups: application to urban raw wastewater.* Journal of virological methods, 2006. **138**(1): p. 131-139.

123. Tartera, C. and J. Jofre, *Bacteriophages active against Bacteroides fragilis in sewage-polluted waters.* Applied and environmental microbiology, 1987. **53**(7): p. 1632-1637.

124. Jofre, J., et al., *Bacteriophages infecting Bacteroides as a marker for microbial source tracking.* Water research, 2014. **55**: p. 1-11.

125. Ebdon, J., M. Muniesa, and H. Taylor, *The application of a recently isolated strain of Bacteroides (GB-124) to identify human sources of faecal pollution in a temperate river catchment.* Water Research, 2007. **41**(16): p. 3683-3690.

126. Ebdon, J.E., et al., *Phages of Bacteroides (GB-124): a novel tool for viral waterborne disease control?* Environmental science & technology, 2011. **46**(2): p. 1163-1169.

127. Vijayavel, K., et al., *Isolation and characterization of Bacteroides host strain HB-73 used to detect sewage specific phages in Hawaii.* water research, 2010. **44**(12): p. 3714-3724.

128. Roslev, P. and A.S. Bukh, *State of the art molecular markers for fecal pollution source tracking in water.* Applied microbiology and biotechnology, 2011. **89**(5): p. 1341-1355.

129. McMinn, B.R., A. Korajkic, and N. Ashbolt, *Evaluation of Bacteroides fragilis GB-124 bacteriophages as novel human-associated faecal indicators in the United States.* Letters in applied microbiology, 2014.

130. Wong, K., et al., *Application of enteric viruses for fecal pollution source tracking in environmental waters.* Environment International, 2012. **45**: p. 151-164.

131. Aw, T.G. and J.B. Rose, *Detection of pathogens in water: from phylochips to qPCR to pyrosequencing.* Current opinion in biotechnology, 2012. **23**(3): p. 422-430.

132. Davison, A., et al., *The DNA sequence of adenovirus type 40.* Journal of molecular biology, 1993. **234**(4): p. 1308-1316.

133. Seif, I., G. Khoury, and R. Dhar, *The genome of human papovavirus BKV.* Cell, 1979. **18**(4): p. 963-977.

134. Kamahora, T., S. Hino, and H. Miyata, *Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells.* Journal of virology, 2000. **74**(21): p. 9980-9986.

135. Takanashi, S., et al., *Characterization of emerging GII. g/GII. 12 noroviruses from a gastroenteritis outbreak in the US in 2010.* Journal of clinical microbiology, 2011: p. JCM. 00305-11.

136. Cantalupo, P.G., et al., *Raw sewage harbors diverse viral populations.* MBio, 2011. **2**(5): p. e00180-11.

137. Bibby, K. and J. Peccia, *Identification of viral pathogen diversity in sewage sludge by metagenome analysis.* Environmental science & technology, 2013. **47**(4): p. 1945-1951.

138. Tamaki, H., et al., *Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate.* Environmental microbiology, 2012. **14**(2): p. 441-452.

139. Rosario, K., et al., *Metagenomic analysis of viruses in reclaimed water.* Environmental microbiology, 2009. **11**(11): p. 2806-2820.

140. Dinsdale, E.A., et al., *Functional metagenomic profiling of nine biomes.* Nature, 2008. **452**(7187): p. 629-632.

141. Shan, T., et al., *The fecal virome of pigs on a high-density farm.* Journal of virology, 2011. **85**(22): p. 11697-11708.

142. Ge, X., et al., *Metagenomic analysis of viruses from the bat fecal samples reveals many novel viruses in insectivorous bats in China.* Journal of virology, 2012: p. JVI. 06671-11.

143. Li, L., et al., *Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses.* Journal of virology, 2010. **84**(14): p. 6955-6965.

144. Li, L., et al., *The fecal viral flora of California sea lions.* Journal of virology, 2011. **85**(19): p. 9909-9917.

145. Phan, T.G., et al., *The fecal viral flora of wild rodents.* PLoS pathogens, 2011. **7**(9): p. e1002218.

146. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-131.

147. Roux, S., et al., *Assessing the diversity and specificity of two freshwater viral communities through metagenomics.* PloS one, 2012. **7**(3): p. e33641.

148. Angly, F.E., et al., *The marine viromes of four oceanic regions.* PLoS biology, 2006. **4**(11): p. e368.

149. Dinsdale, E.A., et al., *Microbial ecology of four coral atolls in the Northern Line Islands.* PloS one, 2008. **3**(2): p. e1584.

150. Desnues, C., et al., *Biodiversity and biogeography of phages in modern stromatolites and thrombolites.* Nature, 2008. **452**(7185): p. 340-343.

151. Meyer, F., et al., *The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes.* BMC bioinformatics, 2008. **9**(1): p. 386.

152. Ye, J., et al., *Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.* BMC bioinformatics, 2012. **13**(1): p. 134.

153. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-227.

154. David, L.A., et al., *Host lifestyle affects human microbiota on daily timescales.* Genome biology, 2014. **15**(7): p. R89.

155. Symonds, E.M. and M. Breitbart, *Affordable Enteric Virus Detection Techniques Are Needed to Support Changing Paradigms in Water Quality Management.* CLEAN–Soil, Air, Water, 2015. **43**(1): p. 8-12.

156. Osuolale, O. and A. Okoh, *Human enteric bacteria and viruses in five wastewater treatment plants in the Eastern Cape, South Africa.* Journal of Infection and Public Health, 2017.

157. Zhou, J., et al., *Source identification of bacterial and viral pathogens and their survival/fading in the process of wastewater treatment, reclamation, and environmental reuse.* World Journal of Microbiology and Biotechnology, 2015. **31**(1): p. 109-120.

158. Qiu, Y., et al., *Assessment of human virus removal during municipal wastewater treatment in Edmonton, Canada.* Journal of applied microbiology, 2015. **119**(6): p. 1729-1739.

159. Blatchley, E.R., et al., *Effects of wastewater disinfection on waterborne bacteria and viruses.* Water environment research, 2007. **79**(1): p. 81-92.

160. Gregory, J.B., R.W. Litaker, and R.T. Noble, *Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of*

enteroviruses in environmental samples. Applied and environmental microbiology, 2006. **72**(6): p. 3960-3967.

161. De Leon, R., et al. *Detection of enteroviruses and hepatitis A virus in environmental samples by gene probes and polymerase chain reaction*. in *Proceedings of the 1990 Water Quality Technology Conference*. 1990. American Water Works Association Denver.

162. ISO, I., *10705–2: Water Quality. Detection and Enumeration of Bacteriophages–Part 2: Enumeration of Somatic. Coliphages.* International Organization for Standardization, Geneva, Switzerland, 2000.

163. EPA, *Method 1602: Male-specific (F+) and Somatic Coliphage in Water by Single Agar Layer (SAL) Procedure*. 2001.

164. Stachler, E. and K. Bibby, *Metagenomic Evaluation of the Highly Abundant Human Gut Bacteriophage CrAssphage for Source Tracking of Human Fecal Pollution.* Environmental Science & Technology Letters, 2014.

165. Minot, S., et al., *Rapid evolution of the human gut virome.* Proceedings of the National Academy of Sciences, 2013. **110**(30): p. 12450-12455.

166. Mizuno, C.M., R. Ghai, and F. Rodriguez-Valera, *Evidence for metaviromic islands in marine phages.* Frontiers in microbiology, 2014. **5**.

167. Sivaganesan, M., et al., *Improved strategies and optimization of calibration models for real-time PCR absolute quantification.* Water research, 2010. **44**(16): p. 4726-4735.

168. Shanks, O.C., et al., *Data acceptance criteria for standardized human-associated fecal source identification quantitative real-time PCR methods.* Applied and environmental microbiology, 2016. **82**(9): p. 2773-2782.

169. Liang, Y., et al., *crAssphage is not associated with diarrhoea and has high genetic diversity.* Epidemiology and Infection, 2016: p. 1-5.

170. Reyes, A., et al., *Going viral: next-generation sequencing applied to phage populations in the human gut.* Nature Reviews Microbiology, 2012. **10**(9): p. 607-617.

171. Minot, S., et al., *Conservation of gene cassettes among diverse viruses of the human gut.* PLoS One, 2012. **7**(8): p. e42342.

172. Carducci, A., et al., *Viral removal by wastewater treatment: monitoring of indicators and pathogens.* Food and Environmental Virology, 2009. **1**(2): p. 85-91.

173. Verhougstraete, M.P., et al., *Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale.* Proceedings of the National Academy of Sciences, 2015. **112**(33): p. 10419-10424.

174. USEPA, *Keeping raw sewage & contaminated stormwater out of the public's water.* 2011.

175. Stachler, E., et al., *Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement.* Environmental Science & Technology, 2017.

176. García-Aljaro, C., et al., *Determination of crAssphage in water samples and applicability for tracking human faecal pollution.* Microbial Biotechnology, 2017.

177. Cinek, O., et al., *Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations.* Journal of medical virology, 2018.

178. Ahmed, W., et al., *Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida.* Water research, 2018. **131**: p. 142-150.

179. Divers, M.T., E.M. Elliott, and D.J. Bain, *Quantification of nitrate sources to an urban stream using dual nitrate isotopes.* Environmental science & technology, 2014. **48**(18): p. 10580-10587.

180. Bain, D.J., et al., *Characterizing a major urban stream restoration project: Nine Mile Run (Pittsburgh, Pennsylvania, USA).* JAWRA Journal of the American Water Resources Association, 2014. **50**(6): p. 1608-1621.

181. Association, N.M.R.W., *State of the Watershed.* 2013.

182. USEPA, E., *Method 1603: Escherichia coli (E. coli) in water by membrane filtration using modified membrane-thermotolerant Escherichia coli agar (modified mTEC).* 2002, EPA 821-R-02-023, US Environmental Protection Agency, Washington [DC].

183. USEPA, E., *Method 1600: Membrane Filter Test Method for Enterococci in Water.* 1997.

184. Divers, M.T., E.M. Elliott, and D.J. Bain, *Constraining nitrogen inputs to urban streams from leaking sewers using inverse modeling: implications for dissolved inorganic nitrogen (DIN) retention in urban environments.* Environmental science & technology, 2013. **47**(4): p. 1816-1823.

185. Agency, U.E.P., *Recreational water quality criteria*. 2012, US Environmental Protection Agency Washington, DC.