REVIEW ARTICLE   OPEN

Metagenomics and the development of viral water quality tools

Kyle Bibby [1], Katherine Crank[1], Justin Greaves[1], Xiang Li[1], Zhenyu Wu[1], Ibrahim A. Hamza[2] and Elyse Stachler[3,4]

Human exposure to pathogenic viruses in environmental waters results in a significant global disease burden. Current microbial water quality monitoring approaches, mainly based on fecal indicator bacteria, insufficiently capture human health impacts posed by pathogenic viruses in water. The emergence of the 'microbiome era' and high-throughput metagenome sequencing has led to the discovery of novel human-associated viruses, including both pathogenic and commensal viruses in the human microbiome. The discovery of novel human-associated viruses is often followed by their detection in wastewater, highlighting the great diversity of human-associated viruses potentially present in the water environment. Novel human-associated viruses provide a rich reservoir to develop viral water quality management tools with diverse applications, such as regulating wastewater reuse and monitoring agricultural and recreational waters. Here, we review the pathway from viral discovery to water quality monitoring tool, and highlight select human-associated viruses identified by metagenomics and subsequently detected in the water environment (namely Bocavirus, Cosavirus, CrAssphage, Klassevirus, and Pepper Mild Mottle Virus). We also discuss research needs to enable the application of recently discovered human-associated viruses in water quality monitoring, including investigating the geographic distribution, environmental fate, and viability of potential indicator viruses. Examples suggest that recently discovered human pathogens are likely to be less abundant in sewage, while other human-associated viruses (e.g., bacteriophages or viruses from food) are more abundant but less human-specific. The improved resolution of human-associated viral diversity enabled by metagenomic tools provides a significant opportunity for improved viral water quality management tools.

INTRODUCTION

Historically, fecal indicator bacteria (FIB) are used to regulate microbiological water quality; however, FIB are not reliable indicators of viruses.[1,2] For example, previous studies have shown a lack of correlation between FIB and gastrointestinal illness in drinking water due to noroviruses,[3,4] and viral outbreaks in recreational waters with FIB at acceptable levels,[5] or in the absence of FIB[6,7] have been documented. Infectious viral pathogens primarily enter the water environment through the release of untreated or inadequately treated wastewater. It is difficult to assess the complete public health and economic impact of waterborne viral infections; however, exposure to fecal-contaminated recreational water is estimated to result in millions of illnesses annually in the United States[8] and hundreds of millions of illnesses globally,[9] and inadequate sanitation and drinking water are predicted to account for greater than 700,000 annual deaths globally.[10] Viruses are also predicted to account for the most significant fraction of human illnesses in sewage-contaminated water under specific exposure scenarios.[11] Improved tools to understand viral pathogens' presence and health impact in sewage-impacted waters are critical to informing engineering and policy efforts to protect human health and improve microbial water quality.

Potential applications of improved viral water quality indicators are broad, including monitoring recreational water and drinking water, ensuring safe wastewater reuse, and agricultural applications (e.g., irrigation water quality and food safety). The ideal viral indicator would have multiple characteristics. Previously proposed criteria include originating from the gut microbiome of target animal species; environmental co-occurrence with pathogens; greater concentration than target pathogens; equivalent or greater environmental resistance to killing stresses as pathogens; specificity to the target host and no environmental sources or replication; correlation to health risk; and, ease in measurement.[12] These criteria should be considered in the selection, development, and evaluation of any potential viral water quality indicator.

Viral water quality indicators have previously been proposed and developed but not yet been widely adopted for regulatory applications. These previously developed indicators mainly fall into two groups: human pathogens and culturable bacteriophages associated with human feces. Examples of human pathogens previously proposed as viral water quality indicators include adenovirus,[13–16] polyomavirus,[13,14,17,18] norovirus,[19] and reovirus.[20] These methods have the advantage of being based upon human viruses capable of causing disease and being highly specific to humans but are challenged by low and variable concentrations in wastewater. Culturable bacteriophages, e.g., bacteriophages infecting E. coli or Bacteroides fragilis,[21] have also been proposed as water quality indicators. These phage-based methods meet many of the above-mentioned criteria for an ideal viral water quality indicator, including higher concentrations than many human viral pathogens in wastewater (somatic and F-specific coliphages previously found to range from 3 to 6 $\log_{10}$ PFU per 100 mL in untreated sewage)[12,22] and more rapid culturability than human viral pathogens.[12] Potential challenges

[1]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN 46556, USA; [2]Department of Water Pollution Research, National Research Centre, Cairo, Egypt; [3]Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA and [4]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland Uberlandstrasse 133, 8600 Dübendorf, Switzerland
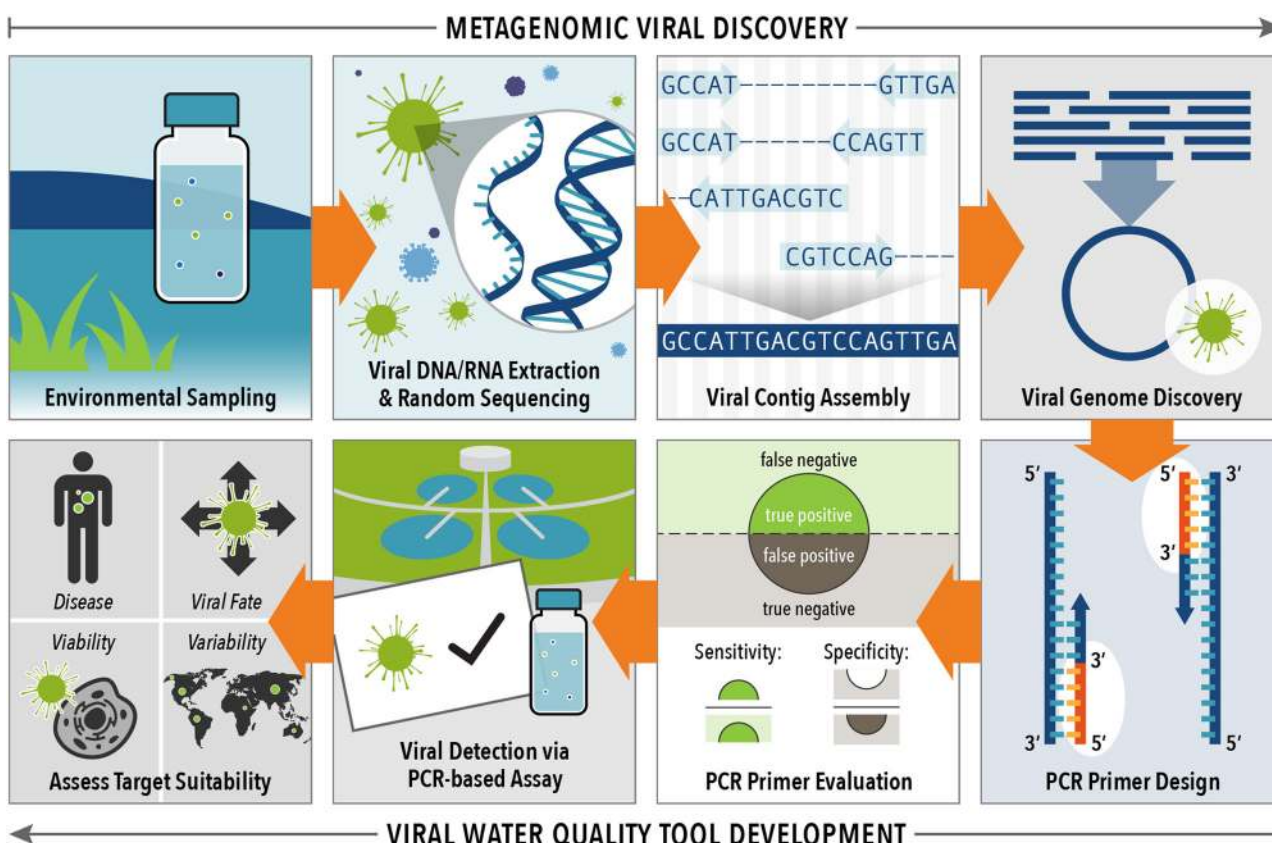Correspondence: Kyle Bibby (KBibby@ND.edu)

Fig. 1 Overview of the pathway from metagenomic viral discovery to viral water quality tool development

in the application of coliphage includes limited specificity to human fecal waste[23] and lower concentrations than other more recently discovered targets, such as those described in this review. Culturable phage-based indicators are currently being evaluated as a regulatory viral water quality indicator by the United States Environmental Protection Agency.[12,24] More recently, alternative viral indicators such as pepper mild mottle virus (PMMoV) and cross-assembly phage (CrAssphage) have been proposed, which will be covered more thoroughly in this review. An opportunity remains to develop improved viral water quality indicators that address shortcomings of current methodologies, namely low concentration in wastewater to facilitate detection or a lack of specificity to human waste.

There has been substantial research interest over the last decade in the 'human microbiome', that is, microbial communities associated with the human body. Technical developments in DNA sequencing have enabled characterization of unculturable microbial communities, with initial discoveries suggesting the relationship between the human microbiome and human health. The microbiome research focus has generated significant insights into the diversity and ecology of the microorganisms that inhabit the human body as well as the discovery of novel human-associated microbiota, including viruses. The discovery of human-associated viruses creates a potential opportunity to develop viral water quality indicators that address current challenges of human specificity and concentration in waste water.

Here, we review current methodologies to move from metagenomic viral discovery to develop a molecular viral water quality indicator. We provide examples of viruses discovered by metagenomics and their subsequent detection in the water environment, with a specific focus on their suitability as a viral water quality indicator. We also discuss some existing shortcomings and research for molecular viral water quality indicators. Finally, we conclude with

an outlook on the potential of developing water quality monitoring tools from newly discovered viruses in the microbiome era.

METAGENOMIC DISCOVERY OF HUMAN-ASSOCIATED VIRUSES

Metagenomic applications for viral discovery[25] and viral pathogen detection[26–28] have been previously reviewed. Here we provide a brief discussion relevant to viral discovery for water quality indicator development. The development of viral metagenomics for direct pathogen detection has shown promise[29,30]; however, it is notable that current untargeted metagenomic methods may miss pathogens detected by targeted methods.[31] In addition, developments to improve the quantitative nature of metagenomics may improve the application of metagenomics for direct viral pathogen detection.[32] The top panel of Fig. 1 shows an overview of metagenomic viral discovery.

Metagenomic sequencing has accelerated viral discovery, including the discovery of uncultured viruses. Currently, only a small fraction of predicted viral genomes exist in available databases;[33] however, recent metagenomic approaches have significantly expanded available viral genomes.[34]

In viral metagenomics, viruses are typically first isolated from a sample, viral DNA or RNA is extracted, and nucleic acids are randomly sequenced (i.e., nucleotide sequence determined) and bioinformatically 'assembled' into longer contiguous sequences (contigs) based upon sequence similarity. Given the relatively short genome size of viruses, contigs of abundant viruses may often be near-complete genomes. Bioinformatic tools, such as VirSorter[35] and MetaVir,[36] have been developed to bioinformatically determine viral genomes from contigs. Following the identification of a potential virus, the genome sequence is often confirmed by re-sequencing using overlapping PCR regions. Potential human-associated pathogens are typically determined

based upon their phylogenetic relationship with previously discovered pathogens.[37,38] Some recently discovered viruses, e.g., crAssphage, share limited sequence similarity with known viruses, challenging their sequence-based detection and assignment of host and gene function.[39]

CHALLENGES DEVELOPING VIRAL WATER QUALITY TOOL FROM NOVEL HUMAN-ASSOCIATED VIRUSES

Traditional virus isolation has been carried out via culturing on a receptive cell line or host. Many viruses discovered by metagenomics are not readily culturable. This is expected; simply put, if these viruses were readily culturable using common existing techniques, these viruses would likely have been previously discovered using culture-based approaches. Novel human-associated viruses are not necessarily 'unculturable', but may require specialized development to culture in the laboratory. The inability to culture these novel viruses hinders culture-based detection in water quality monitoring, essentially requiring molecular (DNA-based or RNA-based) methods, as discussed below. In many cases, the inability to culture novel viruses may also complicate the identification of the viral host or host-range. Similarly, the limited ability to culture these viruses hinders the ability to confirm their disease association via Koch's postulates, leading to the proposal of the 'metagenomic Koch's postulates'.[25] In 'metagenomic Koch's postulates', multiple lines of evidence, primarily co-occurrence with a disease, are used to define a recently discovered microorganism as a causative agent of disease. Similar approaches, including co-occurrence and genomic profiling, can be used to propose viral hosts.

Viruses discovered using metagenomics typically depend on molecular detection in the environment (that is, DNA-based or RNA-based detection) due to the lack of a cultured isolate. Currently, the majority of molecular water quality indicators are PCR-based assays, with quantitative PCR (i.e., real-time or qPCR) as the most common approach. In addition, digital PCR (dPCR) is an emerging water quality monitoring tool.[40] In dPCR a PCR reaction is divided into thousands of individual sub-reactions, and the number of targets in the original sample is quantified based on the number of positive PCR sub-reactions. This development eliminates the need for the standard curve used in qPCR reactions. dPCR has previously been used to detect viruses in natural waters, including norovirus, hepatitis A, and rotavirus.[41,42] When compared to qPCR, assessments have shown that dPCR has greater precision[42] and is less sensitive to PCR inhibitors.[42–44] Results have been more mixed for sensitivity assessments, with some studies showing similar sensitivity as qPCR[41,42] and other studies showing similar or slightly decreased sensitivity.[44] The evidence suggests that dPCR offers several technical advantages over qPCR, with the primary drawbacks being capital costs and the necessity for further technical verification, i.e., limited demonstrations of application for viral monitoring in environmental waters.

Designing PCR assays based on viruses discovered from assembled metagenomic reads comes with specific challenges. The genome assembled from metagenomic sequence data likely represents a consensus genome developed from multiple strains present in the sample. A consensus- or pan-genome results in unknown sequence variation, complicating molecular assay development based on this data. In addition, novel viruses are often assembled from a relatively small number of samples, which may result in inadequate coverage of true viral diversity.

VIRAL WATER QUALITY TOOL DEVELOPMENT

The bottom panel of Fig. 1 shows an overview of the pathway for viral water quality tool development.

Following metagenomic genome or contig assembly, the sequence can be scanned for potential PCR primers that meet
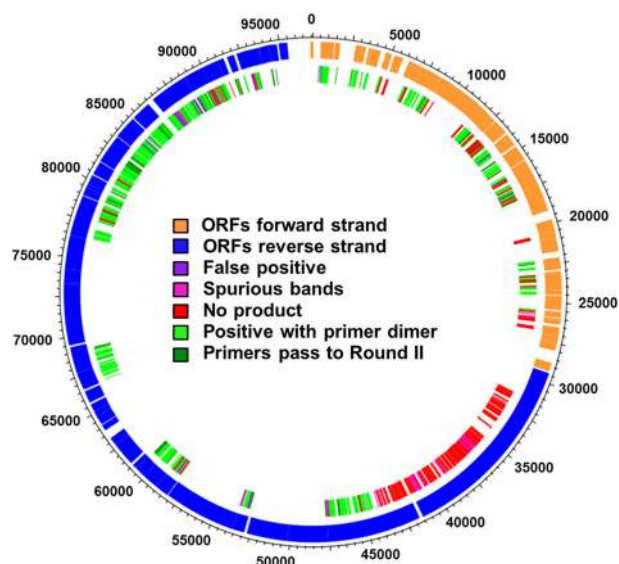

Fig. 2 Genome map demonstrating the shotgun primer design approach used to design a CrAssphage PCR assay by Stachler et al.[49] Genome position is noted by numbers on external ring. The open reading frames of the circular CrAssphage genome are shown in blue and orange on the external ring. Each tick mark in the internal ring represents a primer pair. The outcome of each primer pair screening indicated by legend. Adapted with permission from Stachler et al.[49] Copyright 2017 American Chemical Society

design criteria. A developed PCR assay could then be adapted to qPCR or dPCR chemistry to enable target quantification. In general, PCR primers should be designed from unique, high-quality, and unambiguous DNA regions. Some important rules are: (1) the primer lengths should be from 17 to 25 basepairs, as longer primers may lead to higher annealing temperature, which reduces DNA polymerase activity; (2) GC-content should be between 40 and 60%, and both primers should have similar GC content; (3) the bases should be distributed randomly avoiding more than three identical consecutive bases; (4) the 3' end of primers should have two G or C in last 5 bp; (5) regions of secondary structure, such as hairpins, should be discarded during design; (6) bioinformatics primer design tools, such as NCBI Primer-BLAST, should be used to check for potential false positives after primer design.[45,46] Typically, this process is completed for a single or small number of PCR primer targets, and those assays are subsequently optimized; however, that approach may be challenged by uncertainty in the viral genome composition (for example, unknown sequence diversity) that arises from the metagenomic assembly of the genome. Best practices for PCR assay development and implementation for water quality assessment have previously been detailed.[47] Assays targeting RNA also have some additional considerations, namely the addition of a reverse-transcription step and greater possibility of RNA degradation.[48]

Recently, Stachler et al.[49] took a unique approach to overcome unknown sequence conservation in genomes assembled from metagenomes. In this approach, 384 primer pairs were designed to cover the regions of the CrAssphage genome determined to be suitable for primer design and screened against sewage DNA and composite animal fecal DNA to identify the best performing primer pairs. In the initial evaluation, 18 primer pairs gave false positives in animal fecal DNA, 12 primer pairs produced spurious bands, 123 primer pairs produced no product, 174 primer pairs were positive with primer dimer (suboptimal amplification), and 57 primer pairs passed initial evaluation. An overview of this primer design approach over the CrAssphage genome is shown in Fig. 2. This approach represents a potentially powerful method to

rapidly overcome uncertainty due to unknown sequence variability in genomes determined from metagenomic sequence data.

Two of the primary performance metrics of PCR assays for water quality monitoring are sensitivity and specificity. Sensitivity refers to the percentage of non-spiked target samples correctly identified as positive normalized by the total number of target samples tested (e.g., the percentage of human sewage samples positive for the assay). Specificity refers to the percentage of presumptive negative samples correctly identified as negative normalized by the total number of presumptive negative samples tested (e.g., the percentage of non-human fecal samples negative for the assay). Typically, assays are screened against human and animal-derived fecal samples to assess performance. While there are no formal performance criteria for sensitivity or specificity, assays are often considered to perform well if they demonstrate at least 80% sensitivity and specificity.[50] The specific laboratory, reference samples, and amount of input material may influence reported sensitivity and specificity.[50]

Two additional performance metrics often considered are the limit of detection (LOD) and limit of quantification (LOQ). The LOD is the lowest amount of DNA sample in a given reaction condition that can be statistically detected. LOQ is the lowest quantifiable target concentration above the lower limit of the linear detection range and the assay LOD.[51] Varying definitions of LOD may lead to differing interpretations.[52] These measures must also be updated for varying technologies. For example, the LOD for digital droplet PCR has been previously defined as the [mean of the no-template control measures] + 2.479*[standard deviation of the no-template controls].[51] There is not yet agreed upon metrics for the direct application of semi-quantitative metagenomic approaches.

LOD and LOQ have well-recognized challenges as quantitative descriptors of molecular water quality assays; namely that different sources or amounts of input material significantly alter these measures. Improved descriptors have recently been proposed. The first is the 'equivalent threshold', which is the "quantity of fecal reference material per reaction that corresponds to the LOD of an assay".[52] The second is the 'false positive index', which is the "the difference between the $\log_{10}$ transformed ET values of the target source and a non-target source".[52] Continued development of assay descriptors will improve the quantitative comparison of molecular water quality indicators.

Currently, there is no consensus regarding tested reference library size, reaction input concentration, and assay performance parameters for proposed assays. For instance, in the review from Harwood et al.,[53] over 26 studies and 55 assays were summarized regarding assay performances such as sensitivity, specificity, sample type, LOD and LOQ. From this summary, the diverse reported units for input, LOD, and LOQ values indicate that a standard way to evaluate molecular fecal source tracking methods is needed. Even using the same assay, but in different geographical areas, the parameters varied greatly in both LOD and LOQ, suggesting geographic variability.[52,53] Possible explanations for differential performance between laboratories could be slightly different qPCR reaction conditions, gene copy abundances in different locations, different sampling timing, different DNA extraction techniques, and most importantly, different initial input concentrations as the initial target concentration plays an important role in calculating sensitivity and specificity.[52] Likewise, viral concentration approaches are a critical aspect to assay performance as viral targets may be dilute in target environmental waters; method performance of viral concentration methods has been previously reviewed.[54,55] It is also notable that performance is dependent on the specific PCR assay; for example, a previous evaluation found order of magnitude concentration differences using two different adenovirus qPCR assays.[56] These results demonstrate that assays should be carefully chosen and validated for both different laboratories and geographic locations before use.

EXAMPLES OF METAGENOMIC VIRAL DISCOVERY AND DETECTION IN THE WATER ENVIRONMENT

Recently, there have been several examples of human-associated viruses discovered by metagenomic methods. Below we review some notable examples of virus detection in the water environment following their discovery by metagenomics and the potential to use these viruses as water quality monitoring tools.

Bocavirus

Human bocavirus (HBoV) is a member of the parvovirus family first described in 2005 as an agent of respiratory infection using an early metagenomic approach.[38] HBoV is a non-enveloped virus with single-stranded 5Kb DNA genome. As HBoV is often detected concurrently with established pathogens, the role of HBoV as a causative agent in disease is unresolved due to the lack either of a simplified cell culture system or animal model.[57] HBoV was detected in fecal samples from children who have diarrhea regardless of the presence of respiratory infection, indicating that HBoV may be both an enteric and respiratory pathogen.[58]

HBoV has been detected in wastewater and environmental waters. Table 1 shows a summary of HBoV detection in wastewater samples and the specific assays used. HBoV was detected using qPCR in 100, 81, 79, and 51% of untreated wastewater samples from Egypt, the United States, Italy, and Norway, respectively.[59–62] HBoV was also detected by viral metagenomics in raw sewage from Nepal and the United States.[63] In a 2009 study in Germany, 40% of river water samples contained HBoV at concentrations between 3×10^1 and 2×10^3 genome copies/L.[64] Previous environmental detection of HBoV have employed multiple detection methods, including qPCR[65] and Luminex[66] assays.

The study of 28 children hospitalized during a sewage-contaminated drinking water outbreak in Finland highlighted the potential for environmental spread of HBoV.[67] In three of the four cases multiple viruses were detected; but in one case, HBoV was the only virus detected, suggesting that HBoV was the primary cause of symptoms.[67]

The evidence suggests that the metagenomic discovery of HBoV identified an unrecognized but likely minor source of waterborne viral disease; however, the utility of HBoV as a viral water quality-monitoring tool may be limited by method specificity to wastewater less than 100% in the majority of previous studies.

Cosavirus

Human cosavirus (HCoSV), a genus in the Picornaviridae family, was first identified in 2008 in stool samples of Pakistani children with non-polio acute flaccid paralysis through the use of viral metagenomics.[68,69] HCoSV has a 7.3 Kb single-stranded RNA genome and is genetically diverse.[70] HCoSV was previously detected in stool samples from multiple regions throughout the world.[71,72] The presence of HCoSV strongly correlates with diarrhea;[70] however, HCoSV has also been detected in the stool of healthy individuals.[71] Its higher concentrations in those without gastroenteritis complicates the understanding of its pathogenesis.[72] HCoSV was also previously detected in pig feces, albeit at lower concentrations than found in infected humans.[73]

Table 2 shows a summary of HCoSV detection in wastewater samples and the specific assays used. HCoSV was detected using qPCR in 71 and 25–38% of untreated sewage samples in Japan and the United States, respectively.[59,74,75] HCoSV was also detected by viral metagenomics in raw sewage from Nepal.[63] HCoSV has also been detected using qPCR in treated wastewater samples in the US, France, and Japan, with 1–2 $\log_{10}$ removal following treatment.[74–76] HCoSV was previously detected using qPCR in 29% of river samples in Japan and 12% of tributary

Table 1. Previous studies detecting HBoV in wastewater and associated PCR primers

Location	Number of samples	Percent positive	Method of detection	Citation	Primer name	Sequence, 5′ – 3′	Primer Citation
United States	21	81%	Nested PCR	[59]	BocaF1	CGCCGTGGCTCCTGCTCT	[59]
					BocaR1	TGTTCGCCATCACAAAAGATGTG	
					BocaF2	GGCTCCTGCTCTAGGAAATAAAGAG	
					BocaR2	CCTGCTGTTAGGTCGTTGTTGTATGT	
Norway	102	51%	Real-time PCR	[62]	BoV2391s	GCACAGCCACGTGACGAA	[68]
					BoV2466a	TGGACTCCCTTTTCTTTTGTAGGA	
					BoV2411s-TM	6FAM-TGAGCTCAGGGAATATGAAAGACAAGCATCG-TMR	
Italy	134	79%	Nested PCR	[69]	2028	GAAATGCTTTCTGCTGYTGAAAG	[60]
					2045	GTGGAAATCCCCACACCAGAT	
					2046	GGTGGGTGCTGCCTGGATA	
					2047	TCTTGAATGTCAGTGTCAGACAT	
Egypt	66	100%	qPCR	[61]	HBOV-1	TGGCAGACAACTCATCACAG	[140]
					NP1- R2544	TCTTCGAAGCAGTGCAAGAC	
					Probe-2494	6FAM-ATCATCAGGAACACCC	
						AATCAGCCAC-BHQ-1	
					HBoV234F	GCACTTCCGCATYTCGTCAG	
					HBoV24R	AGCAGAAAAGGCCATAGTGTCA	
					HBoV234F	GCACTTCCGCATYTCGTCAG	
					HBoV3R	GTGGATTGAAAGCCATAATTTGA	

samples in France at lower concentrations than in wastewater samples.[74,76] Notably, HCoSV was recently detected using qPCR at relatively higher concentrations than observed in other environmental waters in the Bagmati River, Nepal, highlighting the importance of geographic variability in marker evaluation.[77]

HCoSV may be a causative agent of waterborne viral disease; however, its variable specificity to wastewater (25–71%) across studies and presence in animal sources (pig) may limit its application for viral water quality monitoring. Notably, the recent detection of elevated HCoSV concentrations in environmental waters in Nepal points to a greater potential for application in specific geographic regions.

Cross-assembly phage

Cross-assembly phage, or "CrAssphage", is a bacteriophage that was discovered in 2014 by metagenomic data mining of human fecal microbiome sequence data.[39] It has a dsDNA, 97Kb genome.[39] The study found that the CrAssphage genome was present in 73% of available human fecal metagenomes, and that CrAssphage was more abundant in the human gut than all other known human gut phages combined.[39] The phage was predicted to be a *Bacteroides* bacteriophage by co-occurrence profiling,[39] and recent successful cultivation of CrAssphage has confirmed this prediction.[78] Recent work has identified CrAssphage as the prototypical virus in a new viral family.[79]

CrAssphage has been successfully detected in both sewage and environmental waters. Table 3 shows a summary of CrAssphage detection in wastewater samples. CrAssphage was detected using metagenomics in 100% of wastewater samples surveyed and was most abundant in samples from North America and Europe and less abundant in samples from Africa and Asia.[80] Subsequently, CrAssphage qPCR assays were developed, and CrAssphage was detected in 100% of untreated wastewater samples from throughout the United States and Spain at average concentrations of greater than 10^9 genome copies/L.[49,81,82] CrAssphage was also

recently detected with 100% occurrence in untreated wastewater samples from Australia.[83] CrAssphage appears to be highly human-associated; however, previous studies have had positive CrAssphage detections in seagull, dog, chicken, cat, and cow fecal samples, albeit at decreased concentrations to those found in human sewage.[49,82,84] Recent studies have also successfully deployed CrAssphage markers to detect fecal pollution in environmental waters.[84,85] Finally, a recent study exploring CrAssphage diversity found CrAssphage in sewage samples globally.[86]

High detection rates and high specificity to human wastewater as well as the high abundance in sewage suggest the significant potential of CrAssphage as a viral water quality indicator and could be a valuable addition to the microbial source tracking toolbox. However, further demonstrations are necessary, including detection in multiple environments, correlation with viral pathogens, and geographic diversity, abundance, and exploration if different specific animal variants of CrAssphage exist.

Klassevirus

Human Klassevirus, also known as Salivirus, was first discovered by metagenomic sequencing of a novel picornavirus in pediatric stool samples in the United States.[37,87] Klassevirus has a 6–8 Kb single-stranded RNA genome. Serological evidence has identified humans as the host for Klassevirus.[88]

Table 4 summarizes Klassevirus detection in wastewater samples. Klassevirus was detected using qPCR in 14.6, 57, and 50% of wastewater samples in the United States, South Korea, and Japan, respectively.[89–91] Notably, using an improved qPCR assay, Klassevirus was detected in 93% of wastewater samples with a maximum concentration of 9.7×10^6 genome copies/L in Japan.[92] Klassevirus was also detected by viral metagenomics in untreated sewage from Nepal.[63] Klassevirus was also detected using qPCR in 44, 29, and 42–50% of treated wastewater samples from France, Japan, and the United States, respectively.[76,89,92] Finally,

K. Bibby et al.

Table 2. Previous studies detecting HCoSV in wastewater and associated PCR primers

Location	Number of samples	Percent positive	Method of detection	Citation	Primer name	Sequence, 5'–3'	Primer Citation
United States	21	38%	Nested PCR	59			
Japan	14	71%	qPCR	74	HCoSV-rtF735-1	TTGTAGYGATGCTGTRTGTGTGTG	72
					HCoSV-rtR827-1	CCAYTGTGGGTCCTTTCG	
					HCoSV-rtP783	FAM-AGCCTCACAGGCCRRAAGCCCTGTC-DDQ1	
United States	24	25%	qPCR	75	Same primers as described above[72]		

Klassevirus was detected using qPCR in nine out of 56 river water samples in Japan (16% detection rate) and two of six samples (33%) in Barcelona river water and in one of six samples (16%) in Rio de Janeiro known to be contaminated with sewage.[92,93]

Klassevirus has been observed globally in wastewater, treated wastewater, and in environmental waters, characteristics necessary for a human fecal pollution indicator. The development of an improved assay by Haramoto et al. that increased detection rates in sewage from ~50 to >90%[92] increases the potential to utilize Klassevirus in water quality monitoring applications; however, the lower concentrations in sewage ($<10^7$ genome copies/L) than other proposed viral indicators may limit its further development.

Pepper mild mottle virus

PMMoV is a plant pathogen that infects a wide variety of pepper cultivars. PMMoV has an approximately 6.3 Kb single-stranded RNA genome.[94] PMMoV was not discovered by viral metagenomics, but is included in this discussion as PMMoV was first proposed as a viral indicator of fecal pollution[95] following the metagenomic observation of the abundance of PMMoV in feces.[96] The presence of PMMoV in feces and wastewater is due to consumption of infected pepper products. PMMoV is mainly specific to human feces but has previously been detected in chicken, geese, seagull and cow fecal samples.[95] The authors posited that this was due to animal consumption of products containing peppers.[95] A recent manuscript reviewed PMMoV in viral water quality monitoring.[97]

Table 5 summarizes of PMMoV detection in wastewater. The original Rosario et al. study found PMMoV using qPCR in 100% of untreated wastewater samples from throughout the United States at concentrations of greater than 10^9 genome copies /L.[95] Subsequent studies using qPCR have found 100% detection rates in untreated sewage in the United States, Costa Rica, and Vietnam.[98,99] An exception to this high detection rate in sewage is a previous study in South Korea with a PMMoV detection rate of 57%.[90] PMMoV was also found in wastewater in Australia[100] and Germany[101].

Studies have also investigated the removal of PMMoV in water and wastewater treatment. Observed PMMoV removal using qPCR in wastewater treatment plants in the United States and Germany varied from <1 to 3.7 $\log_{10}$.[101,102] Limited to no removal was observed in Bolivian wastewater treatment ponds[103] and surface flow wetlands in the United States.[104] For drinking water processes, 1–2 $\log_{10}$ removal was removed by coagulation and filtration in Thailand,[105] and reverse-osmosis removed PMMoV to below the LOD[106]. PMMoV was proposed as a viral indicator in wastewater reuse.[107]

PMMoV has also been detected in environmental waters using qPCR, including 97% of river samples in Germany,[101] 94% of surface water samples and 38% of groundwater samples in Vietnam,[99] 76% of surface water source samples in Japan,[108] and 33% of beach water samples in Australia.[100] Conversely, PMMoV was not detected in any Costa Rican surface water samples despite the detection of FIB and other MST markers, and 100% PMMoV detection in untreated wastewater.[98] Finally, PMMoV was found in 85% of Karst groundwater samples in Mexico (all samples positive for total coliforms)[109] and PMMoV was the most commonly detected virus in groundwater at artificial groundwater recharge site.[110] Observed PMMoV are genetically diverse in the environment.[108]

The abundance of PMMoV in wastewater and high specificity suggests its high potential as a fecal pollution indicator. One potential challenge may be the detection of PMMoV in chicken, seagull, cow, and goose feces, limiting its human-specificity;[97] however, the previously demonstrated specificity is above the widely-accepted 80% threshold. In addition, the high detection rates in environmental waters suggest that PMMoV is persistent,

Table 3. Previous studies detecting CrAssphage in wastewater and associated PCR primers

Location	Number of samples	Percent positive	Method of detection	Citation	Primer name	Sequence, 5′ – 3′	Primer Citation
United States	9	100%	qPCR	[49]	CPQ56-F	CAGAAGTACAAACTCCTAAAAAACGTAGAG	[49]
					CPQ56-R	GATGACCAATAAACAAGCCATTAGC	
					CPQ56-P	FAM- AATAACGATTTACGTGATGTAAC -MGB	
					CPQ64-F	TGTATAGATGCTGCTGCAACTGTACTC	
					CPQ64-R	CGTTGTTTTCATCTTTATCTTGTCCAT	
					CPQ64-P	FAM-CTGAAATTGTTCATAAGCAA-MGB	
Spain	23	100%	qPCR	[81]	crAss-UP	AGGAGAAAGTGAACGTGGAAACA	[81]
					crAss-LP	TAAAGCTTAAAGTTGGTGCTCGTT	
					crAss-probe	FAM-AGGATTTGGAGAAGGAA-MGBNFQ	
United States	8	100%	qPCR	[82]	*Same primers as described above*[49]		
Australia	3	100%	qPCR	[83]	*Same primers as described above*[49]		

Table 4. Previous studies detecting klassevirus in wastewater and associated PCR primers

Location	Number of samples	Percent positive	Method of detection	Citation	Primer name	Sequence, 5′ – 3′	Primer Citation
Japan	14	93%	RT-qPCR	[92]	KLA-F	TCTGCTTGGTGCCAACCTC	[92]
					KLA-R	CCARGCACACACATGAGRGGATAC	
					KLA-TP	FAM-TGCGGGAGTGCTCT-MGB-NFQ	
Japan	28	50%	Nested PCR	[91]	KL2C-F1	CTCGCYGAGGACATCACGGA	[136]
					KL2C-R1	GTACAGGTACACRACCAGTGGCT	
					KL2C-F2	AATCTGCTGCCCAGGCCGC	
					KL2C-R2	AGGGAGATGGCRGAGAGAGCTGT	
					KL3DF	GTCTGGTCTATYGAYTACTCTTGCTTT	
					KL3DR	AGGACGGAGTAGGGRGTRAA	
					KLVPF	GTCACYCCMAACACCTCCACTGAAG	
					KLVPR	TTCTGCRCCATCRGCTCCCGA	
					LG0118	ATGGCAACCCTGTCCCTGAG	[137]
					LG0117	GGAAACCCAACCACGCTGTA	
					LG0119	GCTAACTCTAATGCTGCCACC	
					LG0136	GCTAGGTCAGTGGAAGGATCA	
South Korea	14	57%	Nested PCR	[90]		*Same primers as described above*[92]	
					modKLVPR	CTG GTR GGG GCA GGT TGG AC	[90]
United States	48	14.6%	Nested RT-qPCR	[89]	SAL-L1	CCCTGCAACCATTACGCTTA	[138]
					SAL-R1	CACACCAACCTTACCCCACC	
					SAL-L2	ATTGAGTGGTGCAT(C)GTGTTG	
					SAL-R2	ACAAGCCGGAAGACGACTAC	

potentially limiting its application to detect recent pollution events, although further research is necessary.

RESEARCH NEEDS TO ENABLE MOLECULAR VIRAL WATER QUALITY INDICATORS

As discussed above, molecular methods are the most common approach to deploy novel human-associated viruses as water quality indicators. The application of molecular methods to water quality monitoring has specific challenges that must be addressed to allow viral marker application as a regulatory tool. Below, we discuss some of these challenges, including current approaches and potential pathways forward.

Association of viral water quality indicator with disease following exposure

The ultimate 'gold standard' for application of a viral water quality monitoring target is an association with disease following exposure. In lieu of epidemiological disease association, modeling approaches may be used to associate molecular water quality indicator exposure with human health outcomes. Previous examples demonstrated a statistical framework based upon quantitative microbial risk assessment to associate a molecular sewage pollution indicator with disease outcomes.[11,111] In these models, molecular water quality indicators are used to indicate the volume of wastewater exposure. Typically, a range of

Table 5. Previous studies detecting pepper mild mottle virus in wastewater and associated PCR primers

Location	Number of samples	Percent positive	Method of detection	Citation	Primer name	Sequence, 5′–3′	Primer citation
United States	12	100%	qPCR	[95]	PMMV-FP1	GAG TGG TTT GAC CTT AAC GTT GA	[139]
					PMMV-RP1	TTG TCG GTT GCA ATG CAA GT	
					PMMV-Probe	CCT ACC GAA GCA AAT G	
South Korea	14	57%	RT-PCR	[90]	PMMoV-CP F	AGC GCT CAA GCT CAG ACA C	[90]
					PMMoV-CP R	CAT ATT TCT GAC CCC AAG CAG	
Germany	24	100%	qPCR	[101]	Ha-PMMV1	GTG GCA GCA AAG GTA ATG GT	[101]
					Ha-PMMV2	ATT TGC TTC GGT AGG CCT CT	
Vietnam	2	100%	qPCR	[99]	*Same primers as described above[139]*		
United States	48	100%	qPCR	[102]	*Same primers as described above[139]*		
Costa Rica	8	100%	qPCR	[98]	*Same primers as described above[139]*		

concentrations is used with a Monte-Carlo simulation. For reference, the observed concentration ranges for CrAssphage and PMMoV in sewage have been 7–9.5 and 6–10 $\log_{10}$ genome copies per liter, respectively.[49,95,97,112] These observed concentration ranges tend to be above the ranges for human pathogens such as adenovirus, facilitating their direct detection and use in environmental waters. A statistical approach is then used to correlate the abundance of molecular sewage indicators with select pathogens in sewage to determine pathogen exposure. Finally, a quantitative microbial risk assessment is applied to estimate expected disease outcomes depending on the marker concentration. These models assumed that indicator and pathogen concentrations fall within previously observed ranges and did not consider either indicator or pathogens transport or decay. Future model iterations should be expanded to better include the environmental behavior of both model pathogens and viral indicators. Ultimately an epidemiological association between viral marker exposure and disease outcome will likely be necessary to enable their regulatory application; however, quantitative microbial risk assessment establishes an approach to associate molecular water quality markers with disease and provide a path to bridge viral water quality marker development with disease outcome.

Temporal and geographic variability

The microbial composition in human gut microbiomes varies both temporally[113] and geographically.[114] Similarly, human-associated viruses vary based on individual,[115] age,[116] diet,[115,117,118] and geography.[117,119] Each of these factors could alter the abundance, and thus suitability, of a viral indicator in wastewater regionally and temporally. Indeed, the microbial communities found in sewage vary between cities[120] and the prevalence and distribution of waterborne pathogens; for example, norovirus abundance in sewage varies both regionally and globally,[121] as well as temporally.[122]

Characterization of the regional and global distribution of viral water quality pollution indicators is incomplete. Figure 3 shows the global detections of the viruses highlighted in this review as a demonstration. While the evidence is suggestive of global distribution of these viruses, HBoV, HCoSV, Klassevirus, and PMMoV have each only been detected in sewage from 3.1, 1.5, 2.5, and 4.1% of countries globally. An exception is CrAssphage, which following a recent study exploring CrAssphage diversity[86] has been found in sewage in 39% of countries globally. While there is no available data for the vast majority of countries, it is clear that we do not yet know the distribution of these recently discovered viruses for majority of the world's population or demographics. This limited understanding should be considered

before, and may ultimately limit, the global implementation of these viral water quality tools. Regional, rather than global, application of water quality tools may ultimately be most appropriate and should be locally verified prior to application. It may also be most appropriate to apply a 'toolbox' of indicators to account for geographic and temporal variability. The selection of appropriate local molecular assays may be informed by metagenomic methods.

There is a clear research need to expand our view of viral diversity in sewage, perhaps following the consortium model used by Mayer et al.[123] to survey bacterial fecal pollution markers on six continents. At a minimum, novel viral water quality monitoring tools should be evaluated locally before implementation to assess the potential effect of geographic or temporal variation on assay performance.

Viral fate

The motivation to use a viral-based water quality monitoring tool is that ostensibly it would behave more like pathogenic viruses in the environment than currently used FIB; however, this assumption requires validation. Viral fate characteristics, e.g., decay rate, could also be incorporated into quantitative microbial risk assessment models as described above. In this context viral marker 'fate' includes both persistence and transport in the environment. Viral characteristics, including genome and capsid structure, as well as transmission pathway, drive environmental persistence.[124] One of the most important viral characteristics to consider is genome composition (i.e., DNA- or RNA- genome), as genome composition is a significant predictor of viral fate;[124] however, even within structurally similar viruses, viral persistence can vary by over an order of magnitude.[125]

The inherent variability of viral environmental fate demands multiple demonstrations of co-occurrence as well as comparable persistence and transport between potential viral indicators and pathogenic viruses that the indicators are intended to represent. A recent review identified differing predicted risk based upon the age of sewage contamination, and also that risk may be under-predicted using bacterial fecal indicator markers due to the lesser persistence of these markers compared to viral pathogens.[126] Most critically, viral transport and persistence must be well-characterized in several different environments to enable comparison between viral marker and pathogen fate and subsequent updates to exposure models. These evaluations must also take into account that molecular viral marker transport may differ from viable pathogen transport, as DNA may be more persistent in environmental systems than viable microorganisms as discussed below. Ultimately, it may be most appropriate to employ multiple viral indicators in a 'toolbox' approach to most accurately capture
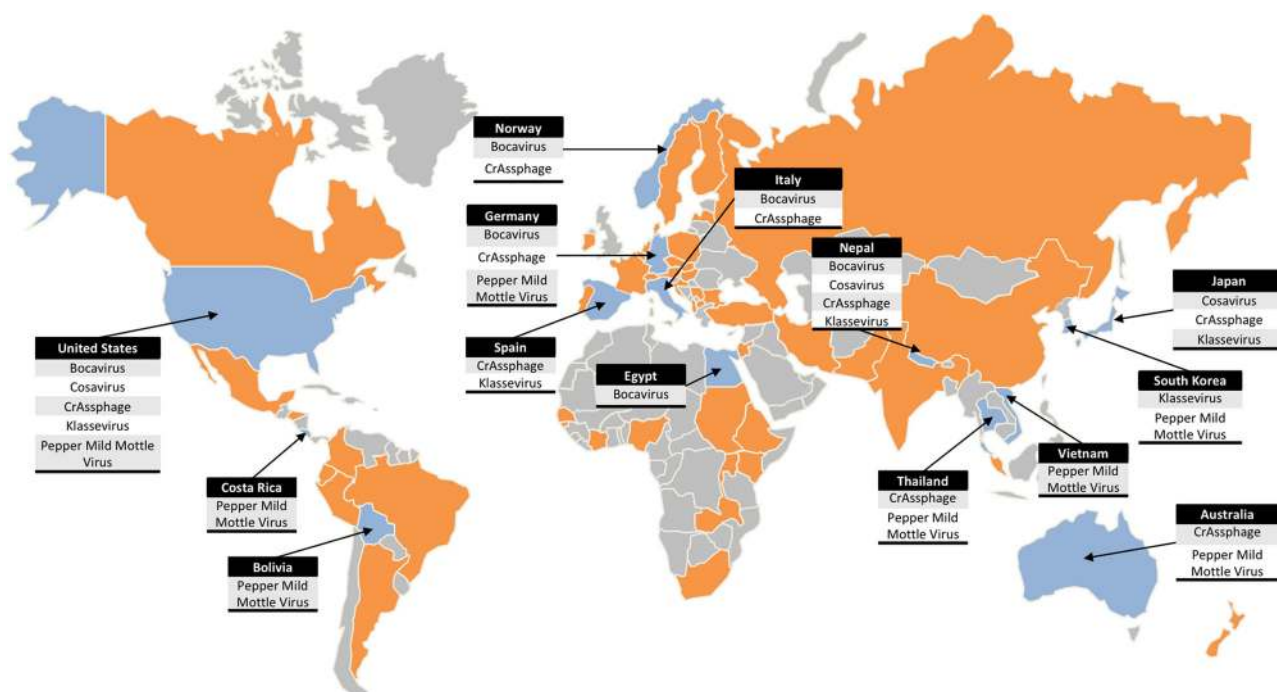
Fig. 3 Countries with HBoV, HCoV, CrAssphage, Klassevirus, or PMMoV identifications in sewage. CrAssphage alone has been documented in orange countries, blue countries contain multiple viruses. Countries not shown are Azerbaijan, El Salvador, Israel, Kazakhstan, Luxembourg, Malta, Singapore and Sri Lanka

the different behavior of pathogenic viruses of interest. Finally, a move from an observational to a mechanistic understanding of viral persistence will better inform the selection of viral surrogates and indicators[127] and provide a framework to integrate the environmental fate of viruses with disease modeling approaches as noted above.

Viral viability

Many viruses discovered via metagenomic methods are not immediately culturable and thus are likely to be developed as molecular detection tools. Molecular methods may detect nucleic acids from both viable and non-viable virus, and nucleic acids may have greater persistence than viable virus.[128,129] Detection of non-viable virus by molecular methods, i.e., 'legacy DNA', may not be informative of the likelihood of infection or a recent contamination event, potentially limiting the application of molecular assays as a water quality monitoring tool. This is especially important in the environment, where a significant fraction of DNA comes from non-viable cells.[130] The question of 'legacy DNA' is critical in many molecular applications and has been recently reviewed;[131] here, we provide a brief discussion relevant to viral water quality monitoring to summarize the state of the field and potential future research directions.

A common approach to discriminate viable virus is through the removal of nucleic acids not contained within the viral capsid. These approaches typically are not capable of excluding virus with damaged capsids that are no longer capable of causing infection but contain the viral genome. The persistence of RNA in the environment is expected to be much less than the persistence of DNA;[132] thus, most approaches are primarily focused on the removal of free DNA. The most common approach currently for removal of legacy DNA is sample digestion with DNAse. Challenges with DNAse include poor removal of particle-bound, non-viable DNA and loss of DNAse activity in some environmental matrices. An alternative approach is the use of intercalating dyes, such as propidium monoazide (PMA). In this approach, intercalating dyes are activated by light and integrate into free DNA,

making the DNA unavailable for PCR amplification or sequencing. The dependence of intercalating dyes on visible light has raised questions about the efficacy of PMA for different environmental matrices and microorganisms.[133]

An alternative approach to discriminating free nucleic acids is via targeting virus capable of successful host binding, demonstrating a functionally intact viral capsid. This approach has previously been demonstrated for norovirus and rotavirus using porcine gastric mucin.[134] Binding assays must be tailored for a specific viral target and have not been demonstrated for widespread environmental application. In addition, binding assays may not capture damage to the viral genome, which may be the predominant mechanism for viral inactivation for certain viruses and conditions.[135]

Finally, advances in viral culturing, for example, the recent successful culturing of norovirus in human intestinal enteroids to determine viral inactivation[131] and the recently reported successful cultivation of CrAssphage,[78] may provide the opportunity to develop culture-based detection approaches for previously unculturable viral targets for viral water quality monitoring. At a minimum, emerging cultivation approaches will be useful to validate methods design to target viable virus as described above.

Additional approaches and validation of methods to discriminate molecular detections from intact virus will be necessary to propel the widespread adoption of molecular tools for water quality monitoring. The path forward will likely require a multi-tiered approach, integrating approaches to exclude free viral nucleic acids and target intact viral capsids, as well as emerging cultivation approaches to confirm the suitability of other methodologies.

SUMMARY AND OUTLOOK

A critical need remains for viral water quality indicators with potentially diverse applications, including recreational and agricultural water quality monitoring, fecal source tracking, and wastewater reuse. The recent rise in human microbiome data and

discovery of human-associated viruses provides a rich opportunity to develop molecular viral water quality monitoring tools. In general, novel human-associated viruses suspected of causing disease, such as HBoV, HCoSV, and Klassevirus in this discussion, are less abundant in wastewater but more human-specific, while bacteriophages such as CrAssphage and food-associated viruses such as PMMoV are more abundant in wastewater but less human-specific. Given that the primary challenge in viral detection is low target abundance and that the potential exists to improve human-specificity of assays through modification or a toolbox approach (that is, using the developed assays in conjunction with others), the immediate richest potential for viral water quality tool development appears to be with bacteriophages or food-associated viruses. Several research areas are necessary to enable the widespread application of these assays, including investigating the association of viral indicators with disease, geographic distribution, and fate and persistence in the environment. In addition, improved tools to exclude non-viable virus from molecular assessments that would not be indicative of a recent pollution event are necessary. There remains a critical need to provide improved viral water quality monitoring tools that would have the potential to reduce hundreds of millions of illnesses globally each year. The improved resolution of human-associated viral diversity enabled by microbiome research provides a significant opportunity for improved viral water quality management tools to address this need.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS
K.B., K.C., J.G. and E.S. performed the initial literature survey and prepared an initial draft of the manuscript. Z.W., X.L. and IAH provided additional revision to the manuscript. K.B. supervised manuscript drafting and revision.

ADDITIONAL INFORMATION

Competing interests: K.B. and E.S. are co-inventors on a patent application entitled "Cross-Assembly Phage DNA Sequences, Primers and Probes for PCR-based Identification of Human Fecal Pollution Sources" (Application Number: 62/386,532). Universities and non-profit researchers interested in using this technology must obtain a research license from the USEPA. To apply for a research license, please request additional information from ftta@epa.gov. The remaining authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. Harwood, V. J. et al. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol.* **71**, 3163–3170 (2005).
2. Payment, P. & Locas, A. Pathogens in water: value and limits of correlation with microbial indicators. *Groundwater* **49**, 4–11 (2011).
3. Strauss, B., King, W., Ley, A. & Hoey, J. R. A prospective study of rural drinking water quality and acute gastrointestinal illness. *BMC Public Health* **1**, 8 (2001).
4. Gruber, J. S., Ercumen, A. & Colford, J. M. Jr Coliform bacteria as indicators of diarrheal risk in household drinking water: systematic review and meta-analysis. *PLoS ONE* **9**, e107429 (2014).
5. Rose, J. B., Mullinax, R. L., Singh, S. N., Yates, M. V. & Gerba, C. P. Occurrence of rotaviruses and enteroviruses in recreational waters of Oak Creek, Arizona. *Water Res.* **21**, 1375–1381 (1987).
6. Hauri, A. et al. An outbreak of viral meningitis associated with a public swimming pond. *Epidemiol. Infect.* **133**, 291–298 (2005).
7. Sinclair, R., Jones, E. & Gerba, C. Viruses in recreational water-borne disease outbreaks: a review. *J. Appl. Microbiol.* **107**, 1769–1780 (2009).
8. Given, S., Pendleton, L. H. & Boehm, A. B. Regional public health cost estimates of contaminated coastal waters: a case study of gastroenteritis at Southern California beaches. *Environ Sci. Technol.* **40**, 4851–4858 (2006).
9. Shuval, H. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *J. Water Health* **1**, 53–64 (2003).
10. Prüss-Ustün, A. et al. Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. *Trop. Med. Int. Health* **19**, 894–905 (2014).
11. Boehm, A. B., Soller, J. A. & Shanks, O. C. Human-associated fecal quantitative polymerase chain reaction measurements and simulated risk of gastrointestinal illness in recreational waters contaminated with raw sewage. *Environ. Sci. Technol. Lett.* **2**, 270–275 (2015).
12. USEPA. *Review of Coliphages as Possible Indicators of Fecal Contamination for Ambient Water Quality* (USEPA, Washington DC, USA, 2015).
13. Albinana-Gimenez, N. et al. Analysis of adenoviruses and polyomaviruses quantified by qPCR as indicators of water quality in source and drinking-water treatment plants. *Water Res.* **43**, 2011–2019 (2009).
14. Hewitt, J., Greening, G. E., Leonard, M. & Lewis, G. D. Evaluation of human adenovirus and human polyomavirus as indicators of human sewage contamination in the aquatic environment. *Water Res.* **47**, 6750–6761 (2013).
15. Rames, E., Roiko, A., Stratton, H. & Macdonald, J. Technical aspects of using human adenovirus as a viral water quality indicator. *Water Res.* **96**, 308–326 (2016).
16. Heim, A., Ebnet, C., Harste, G. & Pring-Åkerblom, P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J. Med. Virol.* **70**, 228–239 (2003).
17. McQuaig, S. M., Scott, T. M., Lukasik, J. O., Paul, J. H. & Harwood, V. J. 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. *Appl. Environ. Microbiol.* **75**, 3379–3388 (2009).
18. Hamza, I. A. et al. Detection of human viruses in rivers of a densly-populated area in Germany using a virus adsorption elution method optimized for PCR analyses. *Water Res.* **43**, 2657–2668 (2009).
19. 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. *Appl. Environ. Microbiol.* **71**, 1870–1875 (2005).
20. Gerba, C. P., Betancourt, W. Q., Kitajima, M. & Rock, C. M. Reducing uncertainty in estimating virus reduction by advanced water treatment processes. *Water Res.* **133**, 282–288 (2018).
21. Ebdon, J., Muniesa, M. & Taylor, H. The application of a recently isolated strain of Bacteroides (GB-124) to identify human sources of faecal pollution in a temperate river catchment. *Water Res.* **41**, 3683–3690 (2007).
22. Rose, J. B. *Reduction of Pathogens, Indicator Bacteria, and Alternative indicators by Wastewater Treatment and Reclamation Processes*. (IWA Publishing, 2005).
23. Harwood, V. J. et al. Performance of viruses and bacteriophages for fecal source determination in a multi-laboratory, comparative study. *Water Res.* **47**, 6929–6943 (2013).
24. Wanjugi, P. et al. Incidence of somatic and F+coliphage in Great Lake Basin recreational waters. *Water Res.* **140**, 200–210 (2018).
25. Mokili, J. L., Rohwer, F. & Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2**, 63–77 (2012).
26. Bibby, K. Metagenomic identification of viral pathogens. *Trends Biotechnol.* **31**, 275–279 (2013).
27. Wong, K., Fong, T.-T., Bibby, K. & Molina, M. Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environ. Int.* **45**, 151–164 (2012).
28. Nieuwenhuijse, D. F. & Koopmans, M. P. Metagenomic sequencing for surveillance of food-and waterborne viral diseases. *Front. Microbiol.* **8**, 230 (2017).
29. Bibby, K., Viau, E. & Peccia, J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Lett. Appl. Microbiol.* **52**, 386–392 (2011).
30. Bibby, K. & Peccia, J. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* **47**, 1945–1951 (2013).
31. Fernandez-Cassi, X. et al. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Sci. Total Environ.* **618**, 870–880 (2018).
32. Nayfach, S. & Pollard, K. S. Toward accurate and quantitative comparative metagenomics. *Cell* **166**, 1103–1116 (2016).
33. Bibby, K. Improved bacteriophage genome data is necessary for integrating viral and bacterial ecology. *Microb. Ecol.* **67**, 242–244 (2014).
34. Paez-Espino, D. et al. Uncovering Earth's virome. *Nature* **536**, 425 (2016).
35. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

36. Roux, S. et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics* **27**, 3074–3075 (2011).

37. Greninger, A. L. et al. The complete genome of klassevirus–a novel picornavirus in pediatric stool. *Virol. J.* **6**, 82 (2009).

38. Allander, T. et al. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc. Natl. Acad. Sci.* **102**, 12891–12896 (2005).

39. Dutilh, B. E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, ncomms5498 (2014).

40. Cao, Y., Raith, M. R. & Griffith, J. F. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water Res.* **70**, 337–349 (2015).

41. Rački, N., Morisset, D., Gutierrez-Aguirre, I. & Ravnikar, M. One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. *Anal. Bioanal. Chem.* **406**, 661–667 (2014).

42. Monteiro, S. & Santos, R. Nanofluidic digital PCR for the quantification of Norovirus for water quality assessment. *PLoS ONE* **12**, e0179985 (2017).

43. Rački, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A. & Ravnikar, M. Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. *Plant Methods* **10**, 42 (2014).

44. Coudray-Meunier, C. et al. A comparative study of digital RT-PCR and RT-qPCR for quantification of Hepatitis A virus and Norovirus in lettuce and water samples. *Int. J. Food Microbiol.* **201**, 17–26 (2015).

45. Rozen, S. & Skaletsky, H. in *Bioinformatics Methods and Protocols* 365–386 (Springer, Totowa, NJ, 2000).

46. Robertson, J. M. & Walsh-Weller, J. in *Forensic DNA Profiling Protocols* 121–154 (Springer, Totowa, NJ 1998).

47. USEPA. *Quality Assurance/quality Control Guidance for Laboratories Performing PCR Analyses on Environmental Samples*. (Office of Water Washington, Washington DC USA, 2004).

48. Girones, R. et al. Molecular detection of pathogens in water–the pros and cons of molecular techniques. *Water Res.* **44**, 4325–4339 (2010).

49. Stachler, E. et al. Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environ. Sci. Technol.* **51**, 9146–9154 (2017).

50. Boehm, A. B. et al. Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* **47**, 6812–6828 (2013).

51. Hindson, C. M. et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **10**, 1003 (2013).

52. Wang, D., Green, H. C., Shanks, O. C. & Boehm, A. B. New performance metrics for quantitative polymerase chain reaction-based microbial source tracking methods. *Environ. Sci. Technol. Lett.* **1**, 20–25 (2013).

53. Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **38**, 1–40 (2014).

54. Haramoto, E. et al. A review on recent progress in the detection methods and prevalence of human enteric viruses in water. *Water Res.* **135**, 168–186 (2018).

55. Hjelmsø, M. H. et al. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS ONE* **12**, e0170199 (2017).

56. Bofill-Mas, S. et al. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. *Appl. Environ. Microbiol.* **72**, 7894–7896 (2006).

57. Schildgen, O. et al. Human bocavirus: passenger or pathogen in acute respiratory tract infections? *Clin. Microbiol. Rev.* **21**, 291–304 (2008).

58. Vicente, D., Cilla, G., Montes, M., Pérez-Yarza, E. G. & Pérez-Trallero, E. Human bocavirus, a respiratory and enteric virus. *Emerg. Infect. Dis.* **13**, 636 (2007).

59. Blinkova, O. et al. Frequent detection of highly diverse variants of cardiovirus, cosavirus, bocavirus, and circovirus in sewage samples collected in the United States. *J. Clin. Microbiol.* **47**, 3507–3513 (2009).

60. Iaconelli, M., Divizia, M., Della Libera, S., Di Bonito, P. & La Rosa, G. Frequent detection and genetic diversity of human bocavirus in urban sewage samples. *Food Environ. Virol.* **8**, 289–295 (2016).

61. Hamza, H., Leifels, M., Wilhelm, M. & Hamza, I. A. Relative abundance of human bocaviruses in urban sewage in Greater Cairo, Egypt. *Food Environ. Virol.* **9**, 304–313 (2017).

62. Myrmel, M., Lange, H. & Rimstad, E. A 1-year quantitative survey of noro-, adeno-, human boca-, and hepatitis E viruses in raw and secondarily treated sewage from two plants in Norway. *Food Environ. Virol.* **7**, 213–223 (2015).

63. Ng, T. F. F. et al. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J. Virol.*, JVl. 00869-00812, (2012).

64. Hamza, I. A., Jurzik, L., Wilhelm, M. & Überla, K. Detection and quantification of human bocavirus in river water. *J. General. Virol.* **90**, 2634–2637 (2009).

65. Pham, N. T. K. et al. A novel RT-multiplex PCR for detection of Aichi virus, human parechovirus, enteroviruses, and human bocavirus among infants and children with acute gastroenteritis. *J. Virol. Methods* **169**, 193–197 (2010).

66. Hamza, I. A., Jurzik, L. & Wilhelm, M. Development of a Luminex assay for the simultaneous detection of human enteric viruses in sewage and river water. *J. Virol. Methods* **204**, 65–72 (2014).

67. Räsänen, S. et al. Mixed viral infections causing acute gastroenteritis in children in a waterborne outbreak. *Epidemiol. Infect.* **138**, 1227–1234 (2010).

68. Yang, Y., Ju, A., Xu, X., Cao, X. & Tao, Y. A novel type of cosavirus from children with nonpolio acute flaccid paralysis. *Virol. J.* **13**, 169 (2016).

69. Kapoor, A. et al. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc. Natl Acad. Sci. USA*, pnas. 0807979105, (2008).

70. Kapusinszky, B., Phan, T. G., Kapoor, A. & Delwart, E. Genetic diversity of the genus Cosavirus in the family Picornaviridae: a new species, recombination, and 26 new genotypes. *PLoS ONE* **7**, e36685 (2012).

71. Ayouni, S. et al. Cosavirus, Salivirus and Bufavirus in Diarrheal Tunisian Infants. *PLoS ONE* **11**, e0162255 (2016).

72. Stöcker, A. et al. Cosavirus infection in persons with and without gastroenteritis, Brazil. *Emerg. Infect. Dis.* **18**, 656 (2012).

73. Okitsu, S. et al. Cosavirus (family Picornaviridae) in pigs in Thailand and Japan. *Arch. Virol.* **161**, 159–163 (2016).

74. Haramoto, E. & Otagiri, M. Occurrence of human cosavirus in wastewater and river water in Japan. *Food Environ. Virol.* **6**, 62–66 (2014).

75. Kitajima, M. et al. Occurrence and genetic diversity of human cosavirus in influent and effluent of wastewater treatment plants in Arizona, United States. *Arch. Virol.* **160**, 1775–1779 (2015).

76. Prevost, B. et al. Large scale survey of enteric viruses in river and waste water underlines the health status of the local population. *Environ. Int.* **79**, 42–50 (2015).

77. Tandukar, S. et al. Presence of Human Enteric Viruses, Protozoa, and Indicators of Pathogens in the Bagmati River, Nepal. *Pathogens* **7**, 38 (2018).

78. Shkoporov, A. N., Khokhlova, E. V., Fitzgerald, C. B., Stockdale, S. R., Draper, L. A., Ross, R. P., & Hill, C. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nature communications **9**, 4781 (2018).

79. Yutin, N. et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol* **3**, 38 (2018).

80. Stachler, E. & Bibby, K. Metagenomic evaluation of the highly abundant human gut bacteriophage CrAssphage for source tracking of human fecal pollution. *Environ. Sci. Technol. Lett.* **1**, 405–409 (2014).

81. García-Aljaro, C., Ballesté, E., Muniesa, M. & Jofre, J. Determination of crAssphage in water samples and applicability for tracking human faecal pollution. *Microb. Biotechnol.* **10**, 1775–1780 (2017).

82. Ahmed, W. et al. Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida. *Water Res.* **131**, 142–150 (2018).

83. Ahmed, W., Payyappat, S., Cassidy, M., Besley, C. & Power, K. Novel crAssphage marker genes ascertain sewage pollution in a recreational lake receiving urban stormwater runoff. *Water Res.* **145**, 769–778 (2018).

84. Ahmed, W., Payyappat, S., Cassidy, M., Besley, C. & Power, K. Novel crAssphage marker genes ascertain sewage pollution in a recreational lake receiving urban stormwater runoff. *Water Res.* **145**, 769–778 (2018).

85. Stachler, E., Akyon, B., Aquino de Carvalho, N., Ference, C. & Bibby, K. Correlation of crAssphage-based qPCR markers with culturable and molecular indicators of human fecal pollution in an impacted urban watershed. *Environ. Sci. Technol.* (2018).

86. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage Robert Edwards, Alejandro Vega, Holly Norman, Maria Cynthia Ohaeri, Kyle Levi, Elizabeth Dinsdale, Ondrej Cinek, Ramy Aziz, Katelyn McNair, Jeremy Barr, Kyle Bibby, Stan Brouns, Adrian Cazares, Patrick A. de Jonge, Christelle Desnues, Samuel Diaz-Munoz, Peter Fineran, Alexander Kurilshikov, Rob Lavigne, Karla Mazankova, David McCarthy, Franklin Nobrega, Alejandro Reyes, German Tapia, Nicole Trefault, Alexander Tyakht, Pablo Vinuesa, Jeroen Wagemans, Alexandra Zhernakova, Frank Aarestrup, Gunduz Ahmadov, Abeer Alassaf, Josefa Anton, Abigail Asangba, Emma Billings, Adrian Cantu, Jane Carlton, Daniel Cazares Lopez, Gyu-Sung Cho, Tess Condeff, Pilar Cortes, Micheal Cranfield, Daniel Cuevas, Rodrigo De la Iglesia, Przemyslaw Decewicz, Michael Doane, Nathaniel Dominy, Lukasz Dziewit, Bashir Elmahi, Murat Eren, Charles Franz, Jingyuan Fu, Cristina Garcia-Aljaro, Elodie Ghedin, Kristen Gulino, John Haggerty, Steven Head, Rene S. Hendriksen, Colin Hill, Heikki Hyoty, Elena Ilina, Mitchell Irwin, Thomas Jeffries, Juan Jofre, Randall Junge, Scott Kelley, Martin Kowalewski, Deepak Kumaresan, Steven Leigh, Eugenia Lisitsyna, Montserrat Llagostera, Julia M. Maritz, Linsey Marr, Angela McCann, Mohammadali Khan Mirzaei, Shahar Molshanski-Mor, Silvia Monteiro, Benjamin Moreira-Grez, Megan Morris, Lawrence Mugisha, Maite Muniesa, Horst Neve, Nam-phuong Nguyen, Olivia Nigro, Anders Nilsson, Taylor O'Connell, Rasha Odeh, Andrew Oliver, Mariana Piuri, Aaron Prussin II, Udi Qimron, Zhe-Xue Quan, Petra Rainetova,

Adan Ramirez-Rojas, Raul Raya, Gillian Rice, Alessandro Rossi, Ricardo Santos, John Shimashita, Elyse Stachler, Lars Stene, Ronan Strain, Rebecca Stumpf, Pedro Torres, Alan Twaddle, Maryann Ugochi Ibekwe, Nicolas Villagra, Stephen Wandro, Bryan White, Andrew Whiteley, Katrine Whiteson, Cisca Wijmenga, Maria Mercedes Zambrano, Henrike Zschach, Bas E. Dutilh bioRxiv 527796; https://doi.org/10.1101/527796

87. Li, L. et al. A novel picornavirus associated with gastroenteritis. *J. Virol.* **83**, 12002–12006 (2009).

88. Greninger, A. L. et al. Serological evidence of human klassevirus infection. *Clin. Vaccine Immunol.* 1584–1588 (2010).

89. Kitajima, M., Iker, B. C., Rachmadi, A. T., Haramoto, E. & Gerba, C. P. Quantification and genetic analysis of salivirus/klassevirus in wastewater in Arizona, USA. *Food Environ. Virol.* **6**, 213–216 (2014).

90. Han, T.-H., Kim, S.-C., Kim, S.-T., Chung, C.-H. & Chung, J.-Y. Detection of norovirus genogroup IV, klassevirus, and pepper mild mottle virus in sewage samples in South Korea. *Arch. Virol.* **159**, 457–463 (2014).

91. Haramoto, E. & Otagiri, M. Prevalence and genetic diversity of klassevirus in wastewater in Japan. *Food Environ. Virol.* **5**, 46–51 (2013).

92. Haramoto, E., Kitajima, M. & Otagiri, M. Development of a reverse transcription-quantitative polymerase chain reaction assay for detection of salivirus/klassevirus. *Appl. Environ. Microbiol.* AEM. 00132-00113, (2013).

93. Calgua, B. et al. Detection and quantification of classic and emerging viruses by skimmed-milk flocculation and PCR in river water from two geographical areas. *Water Res.* **47**, 2797–2810 (2013).

94. Alonso, E. et al. Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistance-breaking tobamovirus in pepper. *J. General. Virol.* **72**, 2875–2884 (1991).

95. Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J. & Breitbart, M. Pepper mild mottle virus as an indicator of fecal pollution. *Appl. Environ. Microbiol.* **75**, 7261–7267 (2009).

96. Breitbart, M. et al. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223 (2003).

97. Symonds, E., Nguyen, K. H., Harwood, V. & Breitbart, M. Pepper mild mottle virus: A plant pathogen with a greater purpose in (waste) water treatment development and public health management. *Water Res.* (2018).

98. Symonds, E. et al. Microbial source tracking in shellfish harvesting waters in the Gulf of Nicoya, Costa Rica. *Water Res.* **111**, 177–184 (2017).

99. Kuroda, K. et al. Pepper mild mottle virus as an indicator and a tracer of fecal pollution in water environments: comparative evaluation with wastewater-tracer pharmaceuticals in Hanoi, Vietnam. *Sci. Total Environ.* **506**, 287–298 (2015).

100. Hughes, B., Beale, D., Dennis, P., Cook, S. & Ahmed, W. Cross-comparison of human wastewater-associated molecular markers in relation to fecal indicator bacteria and enteric viruses in recreational beach waters. *Appl. Environ. Microbiol.* AEM. 00028-00017, (2017).

101. Hamza, I. A., Jurzik, L., Überla, K. & Wilhelm, M. Evaluation of pepper mild mottle virus, human picobirnavirus and Torque teno virus as indicators of fecal contamination in river water. *Water Res.* **45**, 1358–1368 (2011).

102. Schmitz, B. W., Kitajima, M., Campillo, M. E., Gerba, C. P. & Pepper, I. L. Virus reduction during advanced Bardenpho and conventional wastewater treatment processes. *Environ. Sci. Technol.* **50**, 9524–9532 (2016).

103. Symonds, E. et al. A case study of enteric virus removal and insights into the associated risk of water reuse for two wastewater treatment pond systems in Bolivia. *Water Res.* **65**, 257–270 (2014).

104. Rachmadi, A. T., Kitajima, M., Pepper, I. L. & Gerba, C. P. Enteric and indicator virus removal by surface flow wetlands. *Sci. Total Environ.* **542**, 976–982 (2016).

105. Asami, T., Katayama, H., Torrey, J. R., Visvanathan, C. & Furumai, H. Evaluation of virus removal efficiency of coagulation-sedimentation and rapid sand filtration processes in a drinking water treatment plant in Bangkok, Thailand. *Water Res.* **101**, 84–94 (2016).

106. Sangsanont, J., The Dan, D., Thi Viet Nga, T., Katayama, H. & Furumai, H. Detection of pepper mild mottle virus as an indicator for drinking water quality in Hanoi, Vietnam, in large volume of water after household treatment. *J. Environ. Sci. Health, Part A* **51**, 1100–1106 (2016).

107. Sano, D., Amarasiri, M., Hata, A., Watanabe, T. & Katayama, H. Risk management of viral infectious diseases in wastewater reclamation and reuse. *Environ. Int.* **91**, 220–229 (2016).

108. Haramoto, E. et al. Occurrence of pepper mild mottle virus in drinking water sources in Japan. *Appl. Environ. Microbiol.* AEM. 02354-02313, (2013).

109. Rosiles-González, G. et al. Occurrence of pepper mild mottle virus (PMMoV) in groundwater from a karst aquifer system in the Yucatan Peninsula, Mexico. *Food Environ. Virol.* **9**, 487–497 (2017).

110. Betancourt, W. Q. et al. Assessment of virus removal by managed aquifer recharge at three full-scale operations. *J. Environ. Sci. Heal A* **49**, 1685–1692 (2014).

111. Ahmed, W. et al. Quantitative microbial risk assessment of microbial source tracking markers in recreational water contaminated with fresh untreated and secondary treated sewage. *Environ. Int.* **117**, 243–249 (2018).

112. Kitajima, M., Iker, B. C., Pepper, I. L. & Gerba, C. P. Relative abundance and treatment reduction of viruses during wastewater treatment processes—identification of potential viral indicators. *Sci. Total Environ.* **488**, 290–296 (2014).

113. David, L. A. et al. Host lifestyle affects human microbiota on daily timescales. *Genome. Biol.* **15**, R89 (2014).

114. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–+ (2012).

115. Minot, S. et al. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).

116. Lim, E. S., Wang, D. & Holtz, L. R. The Bacterial Microbiome and Virome Milestones of Infant Development. *Trends Microbiol.* **24**, 801–810 (2016).

117. Rampelli, S. et al. Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin. *Environ. Microbiol.* **19**, 4728–4735 (2017).

118. Reyesa, A. et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl Acad. Sci. USA* **112**, 11941–11946 (2015).

119. Holtz, L. R. et al. Geographic variation in the eukaryotic virome of human diarrhea. *Virology* **468**, 556–564 (2014).

120. Shanks, O. C. et al. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl. Environ. Microbiol.* **79**, 2906–2913 (2013).

121. Ahmed, S. M. et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect. Dis.* **14**, 725–730 (2014).

122. Marshall, J. A., Dimitriadis, A. & Wright, P. J. Molecular and epidemiological features of norovirus-associated gastroenteritis outbreaks in Victoria, Australia in 2001. *J. Med. Virol.* **75**, 321–331 (2005).

123. Mayer, R. E. et al. Global distribution of human-associated fecal genetic markers in reference samples from six continents. *Environ. Sci. Technol.* **52**, 5076–5084 (2018).

124. Brainard, J., Pond, K. & Hunter, P. R. Censored regression modeling to predict virus inactivation in wastewaters. *Environ. Sci. Technol.* **51**, 1795–1801 (2017).

125. Aquino de Carvalho, N., Stachler, E. N., Cimabue, N. & Bibby, K. Evaluation of Phi6 persistence and suitability as an enveloped virus surrogate. *Environ. Sci. Technol.* **51**, 8692–8700 (2017).

126. Boehm, A. B., Graham, K. E. & Jennings, W. Can we swim yet? Systematic review, meta-analysis, and risk assessment of aging sewage in surface waters. *Environ. Sci. Technol.* **52**, 9634–9645 (2018).

127. Bibby, K., Aquino de Carvalho, N. & Wigginton, K. In: Research Needs for Wastewater Handling in Virus Outbreak Response, Environmental Science & Technology, ACS Publications, **51**, 2534–2535 (2017).

128. Goetsch, H. E. et al. Fate of the Urinary Tract Virus BK Human Polyomavirus in Source-Separated Urine. *Appl. Environ. Microbiol.* **84**, e02374–17 (2018).

129. Ye, Y. Y., Chang, P. H., Hartert, J. & Wigginton, K. R. Reactivity of Enveloped Virus Genome, Proteins, and Lipids with Free Chlorine and UV254. *Environ. Sci. Technol.* **52**, 7698–7708 (2018).

130. Carini, P. et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol.* **2**, 16242 (2017).

131. Emerson, J. B. et al. Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* **5**, 86 (2017).

132. Tsai, Y. L., Tran, B. & Palmer, C. J. Analysis of Viral-Rna Persistence in Seawater by Reverse-Transcriptase Pcr. *Appl. Environ. Microbiol.* **61**, 363–366 (1995).

133. Fittipaldi, M., Codony, F., Adrados, B., Camper, A. K. & Morato, J. Viable real-time PCR in environmental samples: can all data be interpreted directly? *Microb. Ecol.* **61**, 7–12 (2011).

134. Tian, P., Engelbrektson, A. & Mandrell, R. Two-log increase in sensitivity for detection of norovirus in complex samples by concentration with porcine gastric mucin conjugated to magnetic beads. *Appl. Environ. Microbiol.* **74**, 4271–4276 (2008).

135. Ye, Y., Chang, P. H., Hartert, J. & Wigginton, K. R. Reactivity of enveloped virus genome, proteins, and lipids with free chlorine and UV254. *Environ. Sci. Technol.* **52**, 7698–7708 (2018).

136. Han, T.-H., Kim, C.-H., Chung, J.-Y., Park, S.-H & Hwang, E.-S. Klassevirus infection in children, South Korea. *Emerg. Infect. Dis.* **16**, 1623 (2010).

137. Holtz, L. R. et al. Klassevirus 1, a previously undescribed member of the family Picornaviridae, is globally widespread. *Virol. J.* **6**, 86 (2009).

138. Shan, T. et al. Picornavirus salivirus/klassevirus in children with diarrhea, China. *Emerg. Infect. Dis.* **16**, 1303 (2010).

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

140. Kantola, K. et al. Real-time quantitative PCR detection of four human bocaviruses. *Journal of clinical microbiology* **48**, 4044–4050 (2010).