Research Article

Received: 7 April 2022 Revised: 14 February 2023 Accepted article published: 24 February 2023 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jsfa.12522

Sample and library preparation approaches for the analysis of the virome of irrigation water

Alba Pérez-Cataluña,[a*] Walter Randazzo,[a] Juan F. Martínez-Blanch,[b] Francisco M. Codoñer[b,c] and Gloria Sánchez[a]

Abstract

BACKGROUND: The virome (i.e. community of mainly RNA and DNA eukaryotic viruses and bacteriophages) of waters is yet to be extensively explored. In particular, the virome of waters used for irrigation could therefore potentially carry viral pathogens that can contaminate fresh produce. One problem in obtaining viral sequences from irrigation waters is the relatively low amount of virus particles, as well as the presence of human, bacterial and protozoan cells. The present aimed study was to compare different processing, amplification, and sequencing approaches for virome characterization in irrigation waters.

RESULTS: Our analyses considered percentages of viral reads, values for diversity indices and number of families found in sequencing results. The results obtained suggest that enrichment protocols using two (bezonase and micrococcal nuclease) or four enzymes at once (bezonase, micrococcal nuclease, DNAse and RNase), regardless of an Amicon filtration step, are more appropriate than separated enzymatic treatments for virome characterization in irrigation water. The NetoVIR protocol combined with the ScriptSeq v2 RNA-Seq Library (P0-L20 protocol) showed the highest percentages of RNA viruses and identified the higher number of families.

CONCLUSION: Although virome characterization applied in irrigation waters is an important tool for protecting public health by informing on circulating human and zoonotic infections, optimized and standardized procedures should be followed to reduce the variability of results related to either the sample itself and the downstream bioinformatics analyses. Our results show that virome characterization can be an important tool in the discovery of pathogenic viruses in the environment and can be used to inform and optimize reference-based detection methods provided that appropriate and rigorous controls are included.

Supporting information may be found in the online version of this article.

Keywords: metagenomics; viral enrichment; irrigation water; virome

INTRODUCTION

Water reuse is one option for combating water scarcity and treated wastewater can be an alternative water source for agricultural production. However, wastewater may carry human pathogens, despite the use of water reclamation treatments.[1] For example, it has been shown that water reclamation is not sufficiently efficient for complete viral removal, including that of human enteric viruses.[1,2] Many enteric viruses have been detected in influent and effluent wastewaters, with the most prevalent being adenovirus (AdV), enterovirus (EV), hepatitis A and E viruses (HAV and HEV), norovirus, sapovirus (SaV), astrovirus (HAstV) and rotavirus A (RV).[1,3] Thus, reclaimed water could be a relevant source of potential contamination of produce when used for irrigation purposes. As a result, monitoring treated wastewater for microbiological quality is essential for its intended use. In this sense, the most recent European legislation established a ≥ 6 log decrease of rotavirus, total coliphages or at least one of them (F-specific or somatic coliphages) to validate the efficacy of Wastewater Treatment Plants (WWTPs) in reclaiming water for agricultural irrigation.[4]

Virus detection in environmental waters requires a concentration step given the low amount of viral particles expected.[1] After

* *Correspondence to: Alba Pérez-Cataluña, Department of Preservation and Food Safety Technologies, IATA-CSIC, Valencia, Spain. E-mail: alba.perez@iata.csic.es*

a *Department of Preservation and Food Safety Technologies, IATA-CSIC, Valencia, Spain*

b *ADM-Lifesequencing - Health and Wellness - Adm Nutrition, Valencia, Spain*

c *Danone Nutricia Research, Singapore, Singapore*

sample concentration, traditional methods such as infection on permissive cell lines and/or polymerase chain reaction (PCR) techniques for viral detection and quantification might be applied. However, not every viral pathogen replicates on cell lines and, when the cell system is available, it may be laborious for routine tests, as is the case for human norovirus.[5] Alternatively, PCR techniques stand out for their sensitivity and specificity, they provide results in a short time span and they allow for the detection of unculturable viruses.[1,6] Nevertheless, this methodology is primer-dependent, limiting the detection of viruses whose sequences are not already known. High-throughput sequencing overcomes these restrictions, permitting the study of the entire genetic material present in environmental samples, known as virome when total viral populations are studied.

In the last decade, metagenomics for viral discovery and pathogen detection in clinical and environmental samples have been widely applied.[7] To analyze water samples using viral metagenomics, a virus enrichment step usually precedes the nucleic acid (viral RNA, viral DNA or both) extraction, which are randomly sequenced. The resulting reads are analyzed or further assembled into contigs.[8-16] Despite the relevance of viral populations in environmental waters, viral metagenomic protocols are not yet standardized and the increasing number of sequencing techniques and bioinformatic tools hampers interpretation and comparison of the results. The selective elimination of non-viral targets (e.g. human, bacterial and protozoan cells) from the sample constitutes the first hurdle.[17] For this purpose, several approaches have been evaluated for clinical samples to obtain purified viral nucleic acids in high concentrations by modifying sample preparation and extraction protocols[18,19] or testing viral probes for library enrichment.[1,20] Other studies have focused on the optimization of concentration methods of irrigation water; however, these studies evaluated the efficacy only by real-time PCR.[21,22]

To the best of our knowledge, the present study reports, for the first time, a comprehensive benchmark comparison of different protocols for sample preprocessing and library preparation, with the final goal of providing a workflow for random amplification sequencing to be used for characterizing the virome composition of irrigation waters.

MATERIALS AND METHODS

Virus strains and irrigation water samples

Fecal samples positive for HAstV, norovirus genogroup I (GI) or GII (courtesy of Dr J. Buesa, University of Valencia, Valencia, Spain) and HEV (courtesy of Dr M. J. Alcaraz, Hospital Clínico Universitario, Valencia, Spain) were resuspended (10%, w/v) in phosphate-buffered saline (PBS) containing 2 M $NaNO_3$ (Fisher Scientific, Waltham, MA, USA), 1% beef extract (Conda, Madrid, Spain), and 0.1% Triton X-100 (Fisher Scientific) (pH 7.2), vortexed and centrifuged at 1000 × *g* for 5 min. The supernatants were stored at −80 °C in aliquots.

The HAV cytopathogenic HM-175 strain (ATCC VR-140), the human RV strain Wa (ATCC VR-2018) and the MC0 strain (CECT 100 000) of Mengovirus (MgV) were propagated in FRhK-4 MA-104 and HeLa cell monolayers, respectively. Semipurified stocks were thereafter produced in the same cells by low-speed centrifugations of infected cell lysates (3000 × *g* for 20 min) as reported previously.[23]

Eight water samples were collected at the irrigation head of a commercial greenhouse located in Balsicas (Murcia, Spain). The irrigation water was reclaimed water originating from the effluent streamflow of a wastewater treatment plant. Viruses were concentrated as described previously.[24] Briefly, $MgCl_2$ was added to 200 mL of irrigation water samples to a concentration of 0.05 M, adjusting the pH to 3.5. Samples were filtered through 0.45-µm cellulose nitrate filters (Sartorius, Göttingen, Germany) and then filters were washed out with 5 mL of elution buffer (1% beef extract, 3% Tween-80 and 0.5 M NaCl) and pH adjusted to 9.5. Tubes were shaken for 1 min in a vortex, maintained for 4 min in an ultrasonic bath and shaken again in a horizontal orbital shaker at 250 rpm for 10 min, and then the pH was adjusted to 7. Samples were stored at −80 °C until analysis.

Sample processing and nucleic acid extraction

Aiming to understand how different pretreatments affect the recovery of human enteric virus genomes, one mock sample (referred as '*mock-PBS*') was prepared by artificially inoculating 6.0 log international units (IU) mL^{-1} norovirus GI, 5.7 log IU mL^{-1} norovirus GII, 5.3 log IU mL^{-1} HAV, 6.5 log PCRU mL^{-1} RV, 6.7 log PCRU mL^{-1} MgV, 5.8 log PCRU mL^{-1} HAstrV and 5.2 log IU mL^{-1} HEV in PBS. Additionally, an irrigation water sample was analyzed in parallel for the total virome characterization.

Both samples were processed using five different protocols for the enrichment of viral reads (Fig. 1). First, the NetoVIR protocol (P0) was evaluated with some modifications.[18] Briefly, 500 µL of the sample was homogenized with MP Fast Prep24 5G equipment (MP Biomedicals, Santa Ana, CA, USA) for 40 s at speed 6.0 (as according to the 'Waste Water' program available in the device) and centrifuged at 16000 × *g* for 3 min. Then, 200 µL of supernatant was filtered with 0.8 µm polyethersulfone filters (Sartorius). After filtration, samples were treated with 54 U benzonase (Millipore, Burlington, MA, USA) and 2000 U of micrococcal nuclease (New England Biolabs, Ipswich, MA, USA). Total DNA/RNA was extracted using the NucleoSpin®RNA virus kit (Macherey-Nagel GmbH & Co., Düren, Germany) without carrier RNA. In parallel (Fig. 1), four different enzymatic treatments were evaluated. Protocol 1 (P1) consisted in treating the samples with 52.54 U of benzonase (Millipore) and 2000 U micrococcal nuclease (New England Biolabs) for 2 h at 37°C, then 14 µL of Turbo DNase buffer (Invitrogen, Waltham, MA, USA), 1 U of Turbo DNase (Invitrogen) and 7 U of RNase A (Qiagen, Hilden, Germany) were added and incubated for 1 h at 37°C. To stop enzymatic activities, 28 µL of DNase Inactivation Reagent (Invitrogen) was added to each sample maintained at room temperature and extracted afterwards. Protocol 3 (P3) consisted in applying all the enzymes at once for 2 h at 37°C, and, finally inactivated with 0.5 nM ethylenediaminetetraacetic acid. The protocols 2 (P2) and 4 (P4) included the enzymatic treatments as described for P1 and P3, respectively, followed by a further concentration step with Amicon® Ultra-0.5 filters (Merck KGaA, Darmstadt, Germany). Finally, total RNA and DNA were extracted using the NucleoSpin®RNA virus kit (Macherey-Nagel GmbH & Co.) without adding carrier RNA.

Library preparation and sequencing

Two different library preparation protocols for random sequencing were compared (Fig. 1). One library was generated from 1 to 50 ng of DNA–RNA sample using the Preparation Kit technology (Illumina, San Diego, CA, USA) with slight modifications. In detail, the reverse transcriptase enzyme from the original kit was substituted by Reverse Transcriptase AMV (Thermo Scientific, Waltham, MA, USA), an initial 5-min denaturation step at 95 °C was included, and the number of PCR cycles was increased from 10–15 cycles (as suggested by the

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

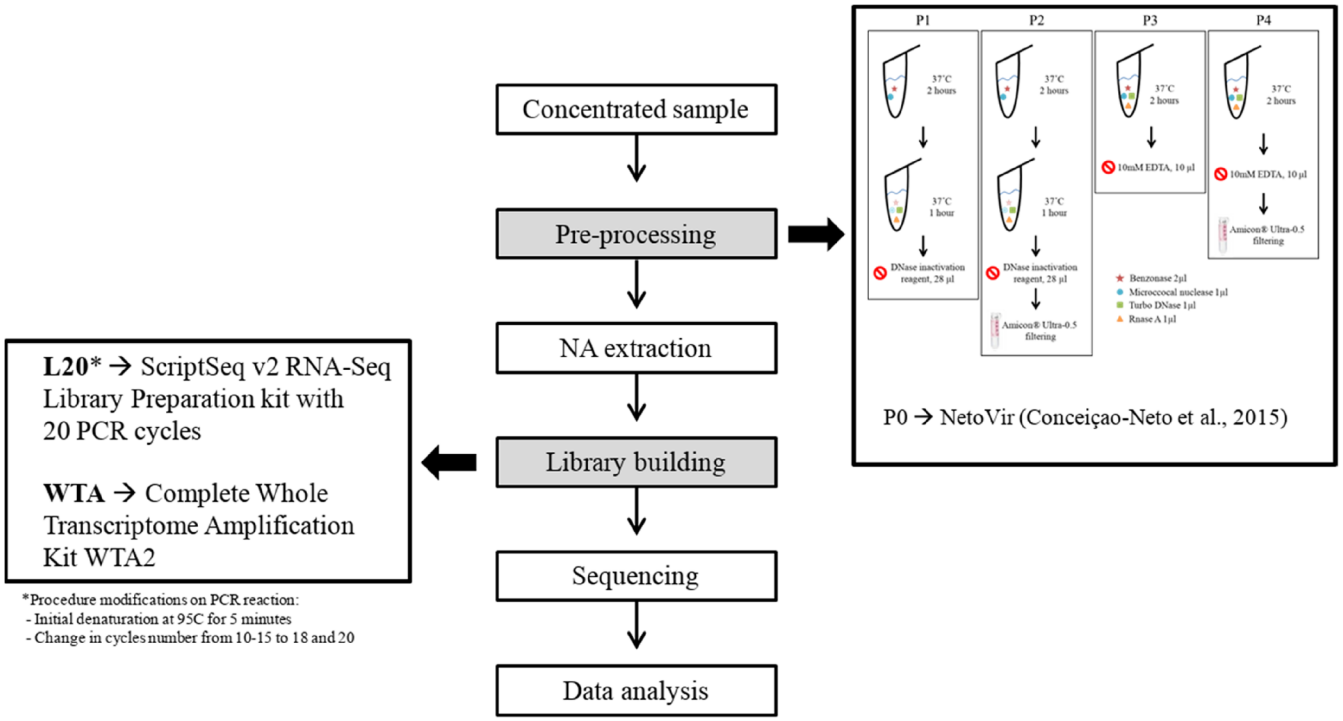


Figure 1. Workflow of the protocols for sample preprocessing and library construction compared in the present study.

manufacturer) up to 20 (L20) to boost the library concentration. A second library (WTA) was prepared using a Complete Whole Transcriptome Amplification kit (Sigma-Aldrich, St Louis, MO, USA) combined with a Nextera XT library preparation kit (Illumina), as described in the NetoVIR protocol.[18] Libraries were normalized, pooled and sequenced using the NextSeq™ 500 system (Illumina), in accordance with the manufacturer's instructions, with a configuration of 150-bp paired-end reads. Library preparation and genome sequencing were performed by Lifesequencing S.L. (Valencia, Spain).

Data analysis

Raw sequences were cleaned for adaptors with *cutadapt*[25] with a minimum overlap length between read and adapter of five nucleotides and a maximum error rate of 0.1. Quality filtering of sequences was carried out with the *reformat.sh* script from BBMap software (https://sourceforge.net/projects/bbmap) using a Phred score of 20 as minimal quality for nucleotides of both ends and 50 bp as minimum length.

Clean reads from mock-PBS samples were aligned to genomes of RV strain Wa (FJ423124-FJ423134), norovirus GI isolate CYY1 (MG049693), MgV isolate M (L22089), HAV strain HM-175 (M14707), HEV genotype 1 (NC_001434.1) and HAstV genotype 1 (NC_001943.1). Alignments were performed using the Burrows-Wheeler Aligner v0.7.17-r1188[26] and the resulting files were indexed by samtools.[27] Genome coverage was calculated with samtools only taking into account nucleotides with at least 20× depth.

Cleaned paired-end reads from water samples were merged in single reads using FLASH, version 1.2.11[28] allowing outies. Merged reads were annotated using BLASTn algorithm[29] with a manually curated in-house database. This database was composed by all the viral sequences (NCBI:txid10239; release September 2020) available at GenBank (https://www.ncbi.nlm.nih.gov/nuccore/?term=viruses%5Borganism%5D). Cut-off values established for the BLASTn analysis were 90% of query sequence coverage and 90% of sequence identity. Rarefaction curves and diversity indices (Shannon and Simpson) were calculated at species level with the R package *vegan* v2.5–6 (https://cran.r-project.org/package=vegan).

Statistical analysis

Normal distribution of data on genome recovery and ecological parameters (i.e. Shannon and Simpson indices) was evaluated with Shapiro–Wilk tests. Significance of the differences in genome coverage for spiked viruses and in ecological values was evaluated using Student's *t*-test for normally distributed data (i.e. for genome coverage methods P0L20, P0WTA, P1L20, P2L20, P4L20, P2WTA and P3WTA, as well as for Shannon and Simpson indices) and Mann–Whitney–Wilcoxon test for not normally distributed data (i.e. methods P3L20, P1WTA and P4WTA in genome coverage results). $P < 0.05$ was considered statistically significant. All the statistical analyses were conducted via R, version 3.6.3 (R Foundation, Vienna, Austria).

RESULTS

Genome recovery and virome composition

Clean reads obtained from a mock-PBS sample were aligned to reference genomes [i.e. RV (FJ423124-FJ423134), norovirus GI (MG049693), MgV (L22089), HAV (M14707), HEV (NC_001434.1) and HAstV (NC_001943.1)], with differences observed according to viral target and sample processing. Specifically, HAV sequences were only found in samples processed with P0, P1-L20, P2-L20 and P2-WTA protocols, whereas norovirus sequences were found in samples prepared according to protocols P1-WTA, P3 and P4-WTA (Fig. 2). HAstrV sequences were not retrieved from any of the assayed protocols. For the other viruses and treatments, the percentage of genome coverage ranged from 0.8% (for MgV

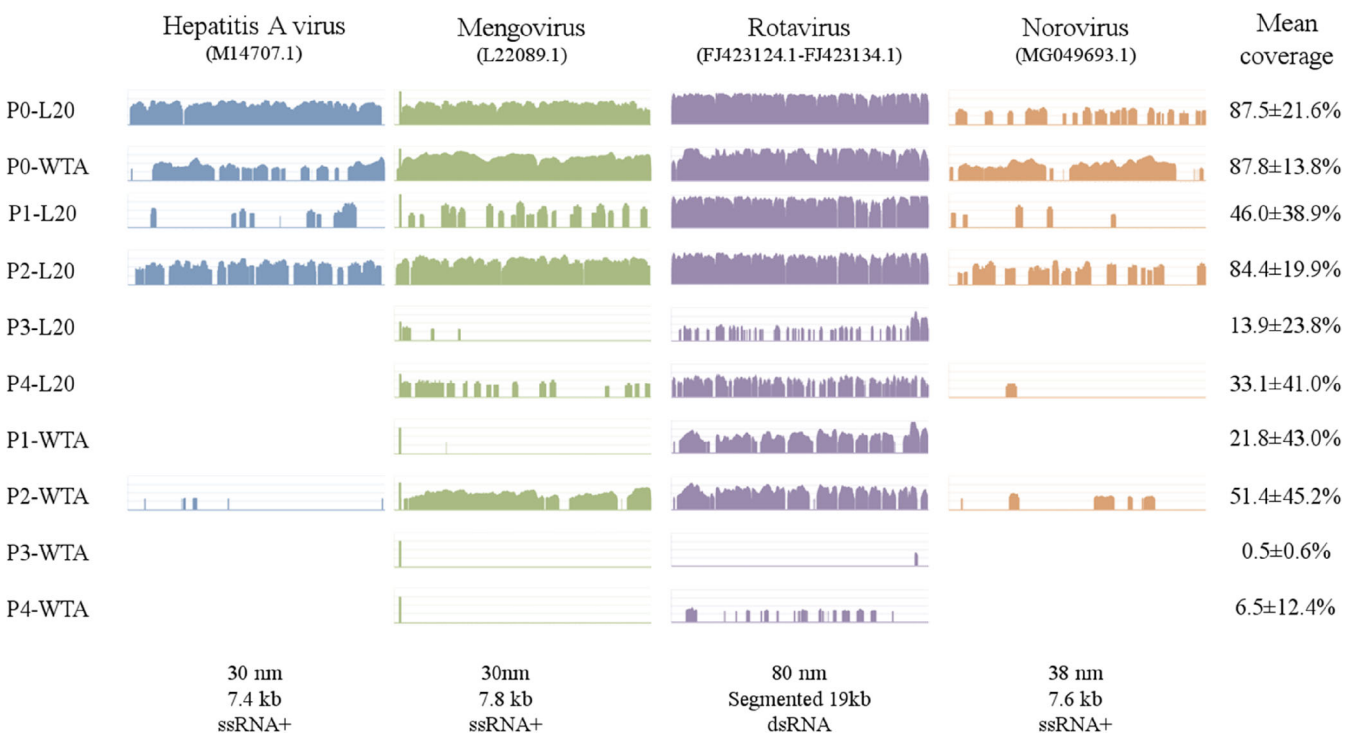
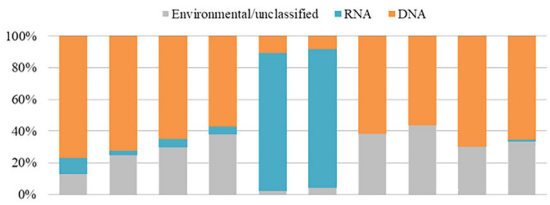


Figure 2. Viral genome coverage of mock-PBS sample analyzed using different protocols. P0, NetoVir; P1, separately enzymatic treatment without filtration; P2, separately enzymatic treatment with filtration; P3, enzymatic treatment at once without filtration; P4, enzymatic treatment at once with filtration; L20, ScriptSeq v2 RNA-Seq Library Preparation Kit; WTA, Complete Whole Transcriptome Amplification kit.

in protocols P3-WTA and P4-WTA) to 99.9% (for RV in protocol P0-L20). Statistical analyses showed significative differences for genome recovery ($P < 0.05$) for P3-L20, P3-WTA and P4-WTA methods compared to P0-L20, P0-WTA and P2-L20, resulting the latter with higher mean genome coverages (87.5%, 87.8% and 84.4%, respectively) (Fig. 2).

To better understand how pretreatment and library preparation affect the entire virome, an irrigation water sample was processed in parallel following the protocols tested for mock-PBS sample. Clean reads were annotated with BLASTn using an in-house database, as described in the Materials and methods. Rarefaction analyses showed that five (i.e. P0-WTA, P1-L20, P3-L20, P4-L20 and P4-WTA) out of 10 sequenced samples reached the plateau, whereas remaining samples were close to stabilization (see Supporting information, Fig. S1). Libraries prepared with WTA kit showed the worst results in rarefaction analyses and the lowest number of identified taxa. The results obtained for each treatment are represented in Fig. 3. Samples P3-L20 and P4-L20 simultaneously showed the highest percentages of viral sequences and the lowest values for both diversity indices. Indeed, five families only were identified among the ones accounting for more than 1% of the viral reads (Fig. 4), with most of them being from the *Reoviridae* family (86.6% and 86.9%, respectively) and corresponding to human rotavirus A sequences. On the other hand, the lowest number of families was obtained in samples where the libraries were built using the WTA kit, as well as the lower percentages of viral reads and low diversity indices, with the exception of sample P4-WTA. Rarefaction analyses supported these results. For those protocols, high percentages of environmental/unclassified viruses were detected, but not RNA viruses.

Characterizing the virome composition at family level, we observed that the number of families representing more than 10% of the total viral reads ranged from 5 (protocols P3-L20 and P4-L20) to 13 (protocols P0-L20) (Fig. 4). The most represented families were *Podoviridae* (25.1 ± 10.2%), *Siphoviridae* (13.1 ± 3.2%) and the 'uncultured virus' taxon (NCBI:txid340016) (27.8 ± 11.3%) for all of the protocols, except for protocols P3-L20 and P4-L20, which mostly identified as *Reoviridae* sequences.



	P0-L20	P0-WTA	P1-L20	P2-L20	P3-L20	P4-L20	P1-WTA	P2-WTA	P3-WTA	P4-WTA
No. reads (Millions)	6.27	6.76	7.74	1.63	6.73	1.75	1.39	2.02	1.48	0.83
No. families	28	22	28	18	33	23	7	8	9	11
% viral seqs	0.13	0.05	0.17	0.17	1	1.46	0.04	0.04	0.04	0.47
Shannon	2.30	2.18	2.02	1.92	0.69	0.66	1.59	1.59	1.72	2.06
Simpson	0.83	0.84	0.79	0.78	0.25	0.24	0.75	0.74	0.77	0.85

Figure 3. Percentage of viral reads obtained for RNA, DNA and environmental or unclassified viruses in the irrigation water sample processed according to the 10 protocols compared in the present study. The number of families, the percentage of viral reads identified, and the Shannon and Simpson indices for each protocol are shown in the lower table. *Number of families that showed more than 10 sequences identified. P0, NetoVir; P1, separately enzymatic treatment without filtration; P2, separately enzymatic treatment with filtration; P3, enzymatic treatment at once without filtration; P4, enzymatic treatment at once with filtration; L20, ScriptSeq v2 RNA-Seq Library Preparation Kit; WTA, Complete Whole Transcriptome Amplification kit.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

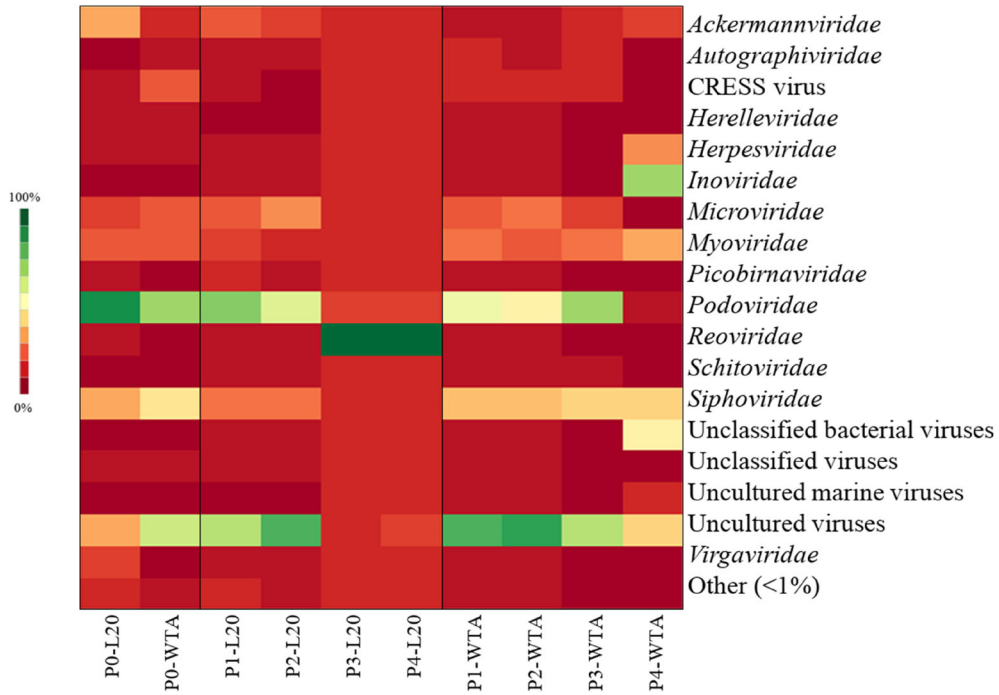
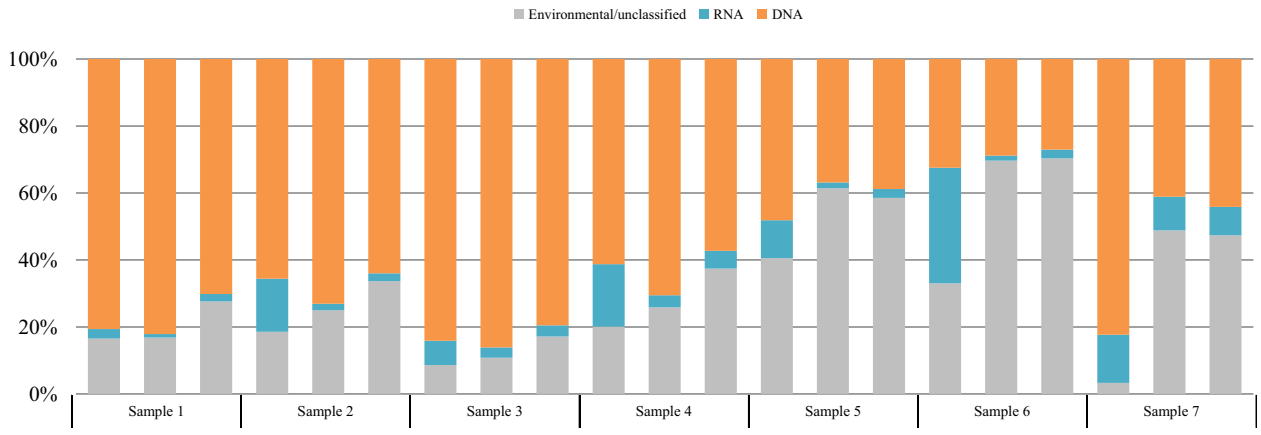


Figure 4. Heat-map showing the percentage of viral families (higher than 1% of the total viral reads) characterizing the irrigation water sample according to the ten protocols tested in the present study. P0, NetoVir; P1, separately enzymatic treatment without filtration; P2, separately enzymatic treatment with filtration; P3, enzymatic treatment at once without filtration; P4, enzymatic treatment at once with filtration; L20, ScriptSeq v2 RNA-Seq Library Preparation Kit; WTA, Complete Whole Transcriptome Amplification kit.

Protocol comparison in irrigation water samples

Based on the preliminary results evaluating ten different protocols, three of them (P0-L20, P1-L20 and P2-L20) were selected for further analysis on seven additional irrigation water samples. These protocols were selected on the basis of the results for the following criteria: (i) distribution of the genomic groups (DNA



	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20	P0-L20	P1-L20	P2-L20
No. Reads (Millions)	3.32	5.56	2.87	2.98	2.41	2.28	5.85	4.90	3.65	3.43	2.19	2.30	4.55	3.97	3.40	5.20	2.35	3.38	5.05	3.72	2.90
No. families*	23	21	19	15	10	12	24	23	21	13	10	12	23	17	17	25	17	20	29	15	17
% viral seqs	0,17%	0,17%	0,19%	0,07%	0,08%	0,08%	0,14%	0,14%	0,14%	0,08%	0,10%	0,11%	0,17%	0,19%	0,21%	0,24%	0,34%	0,36%	0,62%	0,09%	0,10%
Shannon	1,71	1,67	1,83	1,98	1,50	1,63	1,95	1,71	1,86	1,95	1,70	1,76	2,01	1,44	1,57	2,15	1,20	1,25	1,67	1,98	1,89
Simpson	0,69	0,70	0,77	0,81	0,68	0,73	0,73	0,67	0,75	0,81	0,74	0,76	0,78	0,60	0,64	0,83	0,50	0,50	0,72	0,76	0,75

Figure 5. Percentage of viral reads obtained for RNA, DNA, environmental or unclassified viruses in seven irrigation water samples processed with three different protocols. The lower table shows the number of families, the percentage of viral reads identified, and the Shannon and Simpson indices for each sample. *Number of families that showed more than 10 sequences identified. P0, NetoVir; P1, separately enzymatic treatment without filtration; P2, separately enzymatic treatment with filtration; L20, ScriptSeq v2 RNA-Seq Library Preparation Kit.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

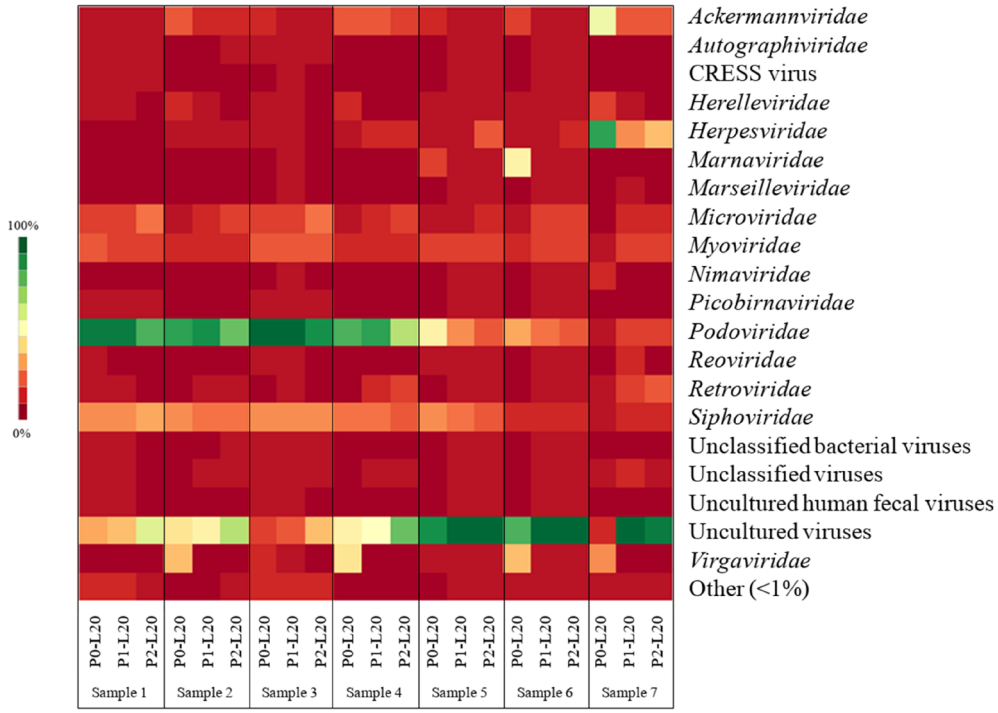


Figure 6. Heat-map showing the percentage of viral families (higher than 1% of the total viral reads) on irrigation water samples treated with the three selected protocols. P0, NetoVir; P1, separately enzymatic treatment without filtration; P2, separately enzymatic treatment with filtration; L20, ScriptSeq v2 RNA-Seq Library Preparation Kit.

and RNA); (ii) diversity indices; (iii) number of families identified; and (iv) recovery of spiked virus sequences in the mock-PBS sample (Figs 2 and 3). Despite the P1-L20 protocol reaching the plateau after rarefaction analysis, P0-L20 and P2-L20 protocols showed rarefaction curves near stabilization and robust results according to selected parameters. The analyses for virome characterization of irrigation water samples have already been presented above and the results are shown in Figs 5 and 6. Analysis of rarefaction curves showed that the majority of the samples were near stabilization (see Supporting information, Fig. S2). Rarefaction analyses were comparable among samples, showing a similar number of identified taxa. Two different sequencing patterns characterized the irrigation water samples: samples with high percentage of DNA viruses (Samples 1–4) and samples mostly characterized by uncultured or unclassified viruses (Samples 5–7), except for Sample 7 treated with protocol P0-L20 in which DNA viruses (belonging to *Herpesviridae* sequences) predominated. Only Sample 6, processed following protocol P0-L20, showed similar percentages for each genomic group (Fig. 5) and, among the RNA viruses, the *Marnaviridae* family was the most represented (Fig. 6). A high abundance of viral signatures belonging to the family *Virgaviridae*, representing plant viruses, was present in four out of seven samples (Samples 2, 4, 6 and 7) processed with the P0-L20 protocol. In general, the P0-L20 protocol showed higher percentages of RNA viruses than other tested protocols. Furthermore, the same protocol allowed a reduction of the percentage of viral reads obtained for environmental or unclassified viruses in irrigation waters.

Methods P0 and P1 differed significantly ($P < 0.05$) based on Shannon and Simpson indices, with method P0 the one showing higher mean values of these indices and therefore greater diversity. The mean number of identified families represented by more than 10 sequences ranged from 12 ± 2 in Sample 4 to 23 ± 2 in Sample 3. The virome profiles were also different within the two groups of samples (Samples 1 to 4 and 5 to 7). Samples 1 to 4 showed 42.9 ± 7.9% of sequences belonging to the *Podoviridae* family, whereas, in Samples 5 to 7, this family represented 9.7 ± 5.5% of sequences. The virome profile of Sample 7 along with protocol P0-L20 differed from the other samples, with the *Herpesviridae* (43.1%) and *Ackermannviridae* (27.0%) families predominating.

DISCUSSION

The scarcity of water resources is forcing producers to use reclaimed wastewater for irrigation, posing some concerns for its microbiological safety. In recent years, metagenomic approaches have been described and applied to monitor the presence of pathogens in water, particularly in wastewater. However, studies comparing viral enrichment protocols on virome characterization of irrigation waters are lacking. In the present study, five different enrichment protocols including filtration and enzyme treatments were tested to evaluate their effect on genome recovery of RNA viruses and on viral composition. Genome recovery analyses showed that the use of protocols P1 to P4 [i.e. enzyme treatments separately (P1 and P2) or jointly (P3 and P4) with (P2 and P4) and without filtering (P1 and P3)] resulted in less recovery of the genomes of the inoculated RNA viruses than with the protocol P0. Although RNase enzyme was added during preprocessing steps to eliminate free and mRNA in the samples, it could have affected the genetic material of capsid-compromised viruses, finally reducing their availability for sequencing.[11] This effect was higher in protocols P3 and P4, in which the four enzymes were added at the same time and incubated for 2 h, increasing the time of action of the RNase enzyme. Furthermore, DNA virus families were present in most of the viromes despite the DNase treatment applied during sample

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

processing, as observed previously.[11] A distinctive pattern could not be retrieved to link the presence of DNA viral signatures to the tested protocols, with considerable variability among the irrigation water samples. The presence of viral DNA reads in all data sets is most likely a result of incomplete digestion of DNA with the Turbo DNase enzyme.

Nevertheless, for the P1 and P2 protocols, differences in genome recovery were observed when the WTA library was used, with this being more noticeable in the case of HAV and norovirus. Differences in protocols P1 and P2 between libraries L20 and WTA could be a result of the use of different retrotranscription enzymes with respect to the process of producing the cDNA combined with the pretreatment with RNase. Regarding virome composition in irrigation water samples, the highest diversity indices were found using the P0 protocol. It is worth noting that viral detection was achieved using the BLASTn database, which is not sensitive for detection of novel viruses, Thus, viral diversity could somehow be restricted to the annotated genome only. This finally implies that the use of the P0-L20 protocol showing the highest viral RNA percentages and the lowest environmental/unclassified viral reads could relevantly inform on the presence of viral pathogens (likely already annotated) in irrigation waters. Furthermore, a reduction in viral diversity when the P1 to P4 protocols are used can be explained by the activity of four enzymes (benzonase, nuclease, DNase and RNase) rather than two enzymes (benzonase and nuclease) used in P0. As occurred for other matrices, the use of enzymes, as well as filtration steps, reduces the number of reads and species detected.[30]

With regard to potential human-pathogenic viruses, the families detected in previous metagenomics studies of sewage include *Astroviridae*, *Caliciviridae*, *Picobirnaviridae*, *Picornaviridae* and *Reoviridae*.[8,12,31,32] In the present study, we recovered *Picobirnaviridae* and *Reoviridae* viromes in irrigation water. Specifically, the high relative abundance of picobirnaviruses warrants further investigation into their use as water quality markers via virome characterization methods, as previously proposed by Adriaenssens *et al.*[11] We cannot rule out the possibility that other pathogenic RNA viruses were present but went undetected by our protocol. The expected low concentration in irrigation water constitutes a bottleneck that needs to be overcome by further optimization of wet-lab procedures and bioinformatics analyses. Specifically, it would be of interest to describe human pathogenic viral signatures at the species level to better understand the dissemination in the environment and water cycle, with relevant implications for public health.

CONCLUSIONS

The results obtained in the present study show that the use of protocols involving two enzymes for virus enrichment (P0-L20) and those that use a treatment with four enzymes in two differentiated steps, with (P2-L20) or without (P1-L20) filtration, are useful for the study of the virome present in irrigation waters. Although there were differences between the samples, these could be a result of the nature of the samples themselves. These data, together with the variability of the virome observed depending on the sequencing method used, make it clear that there is not yet a completely optimized method for virome studies in irrigation waters and that additional studies are necessary to better characterize the total virome circulating in irrigation water.

ACKNOWLEDGEMENTS

This study was supported by the 'VIRIDIANA' project AGL2017-82909 (AEI/FEDER, UE) funded by Spanish Ministry of Science, Innovation and Universities; the 'MAGIC' project from the Center of Produce Safety (CPS, no. 2018CPS10); and the APOTI grant (APOTIP/2018/007) from the Generalitat Valenciana. IATA-CSIC is a Centre of Excellence Severo Ochoa (CEX2021-001189-S MCIN/AEI/10.13039/501100011033). AP-C is recipient of a postdoctoral grant from the Generalitat Valenciana (APOSTD/2021/292). Names of specific vendors, manufacturers or products are included for informational purposes only and does not imply endorsement by authors or their affiliations.

AUTHOR CONTRIBUTIONS

GS, WR and AP-C designed the work. WR and AP-C performed the nucleic acid extractions. AP-C performed the bioinformatics and data analysis. JM-B and FC performed the library preparations and sample sequencing. GS was responsible for funding acquisition. All of the authors wrote and reviewed the manuscript submitted for publication.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

REFERENCES

1 Haramoto E, Kitajima M, Hata A, Torrey JR, Masago Y, Sano D *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).
2 Sano D, Amarasiri M, Hata A, Watanabe T and Katayama H, Risk management of viral infectious diseases in wastewater reclamation and reuse: review. *Environ Int* **91**:220–229 (2016).
3 Qiu Y, Lee BE, Neumann N, Ashbolt N, Craik S, Maal-Bared R *et al.*, Assessment of human virus removal during municipal wastewater treatment in Edmonton, Canada. *J Appl Microbiol* **119**:1729–1739 (2015).
4 Parliament E. Regulation (EU) 2020/741 of the European Parliament and of the Council of 25 May 2020 on minimum requirements for water reuse. 2020.
5 Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR *et al.*, Replication of human noroviruses in stem cell-derived human enteroids. *Science (80-).* **353**:1387–1393 (2016).
6 Nieuwenhuijse DF and Koopmans MPG, Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Front Microbiol* **8**:8(FEB) (2017).
7 Hamza IA and Bibby K, Critical issues in application of molecular methods to environmental virology. *J Virol Methods* **266**:11–24 (2019 Apr).
8 Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP *et al.*, Raw sewage harbors diverse viral populations. *MBio* **2**:e00180-11 (2011).
9 Aw TG, Wengert S and Rose JB, Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses. *Int J Food Microbiol* **223**:50–56 (2016).
10 Prevost B, Lucas FS, Ambert-Balay K, Pothier P, Moulin L and Wurtzer S, Deciphering the diversities of astroviruses and noroviruses in wastewater treatment plant effluents by a high-throughput sequencing method. *Appl Environ Microbiol* **81**:7215–7222 (2015).
11 Adriaenssens EM, Farkas K, Harrison C, Jones DL, Allison HE and McCarthy AJ, Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. *mSystems* **3**:e00025-18 (2018).
12 Fernandez-Cassi X, Timoneda N, Martínez-Puchol S, Rusiñol M, Rodriguez-Manzano J, Figuerola N *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).
13 Gu X, Tay QXM, Te SH, Saeidi N, Goh SG, Kushmaro A *et al*., Geospatial distribution of viromes in tropical freshwater ecosystems. *Water Res* **137**:220–232 (2018 Jun).
14 Guerrero-Latorre L, Romero B, Bonifaz E, Timoneda N, Rusiñol M, Girones R *et al*., Quito's virome: metagenomic analysis of viral diversity in urban streams of Ecuador's capital city. *Sci Total Environ* **645**: 1334–1343 (2018 Dec).
15 Wang Y, Jiang X, Liu L, Li B and Zhang T, High-resolution temporal and spatial patterns of Virome in wastewater treatment systems. *Environ Sci Technol* **52**:10337–10346 (2018 Sep).
16 Costeira R, Doherty R, Allen CCR, Larkin MJ and Kulakov LA, Analysis of viral and bacterial communities in groundwater associated with contaminated land. *Sci Total Environ* **656**:1413–1426 (2019 Mar).
17 Reyes A, Semenkovich NP, Whiteson K, Rohwer F and Gordon JI, Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol* **10**:607–617 (2012 Sep).
18 Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W *et al*., Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep* **5**:5 (2015).
19 Cotten M, Oude Munnink B, Canuti M, Deijs M, Watson SJ, Kellam P *et al*., Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS One* **9**:e93269 (2014).
20 Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ *et al*., Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* **6**:e01491–e01415 (2015).
21 Albinana-Gimenez N, Clemente-Casares P, Calgua B, Huguet JM, Courtois S and Girones R, Comparison of methods for concentrating human adenoviruses, polyomavirus JC and noroviruses in source waters and drinking water using quantitative PCR. *J Virol Methods* **158**:104–109 (2009 Jun).
22 De Keuckelaere A, Baert L, Duarte A, Stals A and Uyttendaele M, Evaluation of viral concentration methods from irrigation and processing water. *J Virol Methods* **187**:294–303 (2013 Feb).
23 Falcó I, Randazzo W, Gómez-Mascaraque LG, Aznar R, López-Rubio A and Sánchez G, Fostering the antiviral activity of green tea extract for sanitizing purposes through controlled storage conditions. *Food Control* **84**:485–492 (2018).
24 López-Gálvez F, Truchado P, Sánchez G, Aznar R, Gil MI and Allende A, Occurrence of enteric viruses in reclaimed and surface irrigation water: relationship with microbiological and physicochemical indicators. *Int J Food Microbiol* **121**:1180–1188 (2016).
25 Martin M, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10-12 Next Gener Seq Data Anal - 1014806/ej171200. 2011 May 2.
26 Li H and Durbin R, Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**:1754–1760 (2009 Jul).
27 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al*., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079 (2009 Aug).
28 Magoč T and Salzberg SL, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**:2957–2963 (2011 Nov).
29 Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N *et al*., BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* **41**:W29–W33 (2013 Jul).
30 Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S *et al*., Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods* **213**:139–146 (2015 Mar).
31 Bibby K and Peccia J, Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* **47**: 1945–1951 (2013 Feb).
32 Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L *et al*., High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* **86**:12161–12175 (2012 Nov).

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.