CENTER for PRODUCE SAFETY

## CPS 2017 RFP
## FINAL PROJECT REPORT

### Project Title
Metagenomics to identify viral indicators in the produce chain

### Project Period
January 1, 2018 – December 31, 2019 (extended to February 28, 2020)


### Principal Investigator
Gloria Sánchez Moragas
IATA-CSIC
46980 Paterna, Spain
T: +34-963-900-022
E: gloriasanchez@iata.csic.es

### Co-Principal Investigator
Jesús Rodríguez Díaz
University of Valencia
Department of Microbiology and Ecology
46010 Valencia, Spain
T: +34-963-983-316
E: jesus.rodriguez@uv.es


---

### Objectives
1. *Optimization of sample preparation procedure for viral metagenomics from irrigation water samples.*
   - *1.1. Optimization of the sample preparation by using spiked samples*
   - *1.2. Validation of the optimized procedure in positive water samples*
2. *Determination of the viral community composition of samples previously analyzed that tested positive or negative for the presence of human pathogenic viruses.*
   - *2.1. Determination of the entire viriome of archived samples*
   - *2.2. Collection of irrigation waters and produce samples*
   - *2.3. Determination of the entire viriome of newly collected samples*
3. *Identification of specific viral species or groups whose presence/abundance correlates with the occurrence of human pathogenic viruses in stools, irrigation waters and produce.*


### Funding for this project provided by the Center for Produce Safety through:
CPS Campaign for Research

## FINAL REPORT

## Abstract

The reuse of water, including for irrigation, cooling, and other non-potable applications, is an emerging topic due to climate change and water scarcity. Irrigation water, fecal-contaminated surfaces, and workers are potential transmission sources of human enteric viruses to produce, so ensuring irrigation water quality and minimizing human-fecal contamination are critical to limiting produce-associated viral diseases. Fecal bacteria have traditionally been used as an indicator for the presence of pathogenic microorganisms, but these indicators fail to detect the presence of human pathogenic enteric viruses due to their higher environmental resistance.
Our project used state-of-the-art metagenomics approaches to study the wide diversity of viruses, i.e., the virome, which is a diverse community of mainly eukaryotic RNA and DNA viruses and bacteriophages, to identify a suitable viral indicator/s in irrigation water, produce and fecal samples. We developed an optimized procedure to concentrate viruses from irrigation water, produce and stool samples and a pipeline for the sequencing of RNA and DNA viruses. Viral particles were enriched from sample concentrates using a filtration and nuclease digestion procedure prior to total nucleic acid extraction and preparation of sequencing libraries for each individual sample. Additionally, samples were also tested for the occurrence of different enteric viruses (norovirus GI, norovirus GII, hepatitis A virus, rotavirus and astrovirus) using RT-qPCR (reverse transcription–quantitative polymerase chain reaction). Overall, we have determined the viral community composition of 254 samples that tested positive or negative for the presence of human pathogenic viruses. After virome profile comparison between positive and negative samples for enteric viruses detection, our results suggest cross-assembly phage (crAssphage) as a potential indicator to assess the presence of human enteric viruses in irrigation water, stool and produce samples, although 100% correlation was not achieved. Moreover, the ideal indicator should be easily detected by simple, inexpensive laboratory testing in the shortest time with accurate results. The detection of crAssphage is performed by molecular methods so far, thus not inferring viral infectivity. Additionally, a few more phages of potential interest have been identified and further studies are needed to assess their suitability to predict virus presence.

## Background

The human health risks associated with consumption of produce containing human enteric viruses are well recognized, with numerous foodborne outbreaks documented and frequently linked with food-handling issues or presence of human viruses in irrigation water samples (Machado-Moreira et al., 2019). A wide variety of viruses may be transmitted by food, nevertheless the most frequently reported viruses are human noroviruses, causing 120 million cases of gastroenteritis each year, and hepatitis A virus (HAV). In the United States, norovirus is responsible for most produce-related outbreaks, followed by *Salmonella* (Bennett et al., 2018).

Irrigation water is one of the vehicles identified for the contamination of produce with pathogenic microorganisms, including human pathogenic viruses, during primary production. Several norovirus and HAV viral outbreaks have been ascribed to crops contaminated in the field, suggesting contamination by irrigation (Dentinger et al., 2001; Ethelberg et al., 2010; Falkenhorst et al., 2005). The presence of human enteric viruses in irrigation waters has been extensively documented (Li et al., 2018; López-Gálvez et al., 2016; Randazzo et al., 2016; Tian et al., 2017). Among others, the viruses most commonly detected in irrigation waters include human norovirus, astrovirus (HAstV), rotavirus A (RV), and HAV (Ashbolt, 2015).

Moreover, several publications have reported the presence of human pathogenic viruses in produce. Belgium, Canada and France have provided data on the prevalence of norovirus in more than 1,000 samples of fresh produce, where norovirus genomes were frequently detected (28.2, 33.3 and 50%, in samples from Canada, Belgium and France, respectively) (Baert et al., 2011). Human pathogenic viruses have also been detected in market produce (Cheong et al., 2009; El-Senousy et al., 2013; Hernández et al., 1997; Monge and Arias, 1996). In salad vegetables from European countries, samples were positive in 1.32, 3.42, 2 and 2.95% for HAV, hepatitis E virus (HEV), norovirus GI and GII, respectively (Kokkinos et al., 2012). In Mexican produce samples, HAV was found in 28.2% of the samples, norovirus in 32.6% and rotavirus in 13.0% (Felix-Valenzuela et al., 2012). Recently, norovirus prevalence in Australian leafy greens at retail was reported at 2.2% (Torok et al., 2019).

Human pathogenic viruses can get on edible parts of the plants not only by direct contact with irrigation water or feces but also by their internalization through the roots (Dicaprio et al., 2012; Chandrasekaran and Jiang, 2018). Microbiological safety of irrigation water and produce is currently assessed through the use of bacterial indicators. Despite the universal use of coliforms and *Escherichia coli* as indicators to predict the risk of exposure to pathogens of fecal origin in water and produce, bacteria have been shown to be poor indicators of human enteric viral contamination (reviewed by Jofre et al., 2016). Structurally viruses are diverse and are quite distinct from bacterial cells. They also display significantly different resistance and susceptibility responses to environmental conditions such as UV irradiation, and water and sewage treatment processes.

Bacteriophages (including F-specific RNA, somatic coliphages and Bacteroides spp. phages) are generally better predictors due to similarities in morphology and survival dynamics. In addition to their value as indicators of fecal contamination, indicator bacteriophages have been viewed as potential surrogates of human pathogenic viruses for inactivation studies. However, correlations between the presence and levels of bacteriophages and human enteric viruses in water have been studied with disparate results (Jofre et al., 2016). Moreover, other viral indicators have also received significant attention, and adenovirus, human polyomavirus, cross-assembly phage (crAssphage) and pepper mild mottle virus (PMMoV) have been proposed as indicators of human pathogenic viruses (McQuaig et al., 2012; Rosario et al., 2009; Rusiñol et al., 2014).

Main goal of the project:

The goal of the study was to use next-generation sequencing (NGS) to characterize the virome (a diverse community of mainly eukaryotic RNA and DNA viruses and bacteriophages) from archived and newly collected irrigation water samples, produce and feces in order to identify a more significant indicator/s of the presence or absence of human pathogenic viruses.

## Research Methods and Results

During the current project, a total of 254 samples—192 irrigation water, 40 produce and 22 stool samples—were analyzed. Irrigation water and produce samples were collected from two Mediterranean fresh produce growing regions in Spain (i.e., Valencia and Murcia).

**Objective 1:** *Optimization of sample preparation procedure for viral metagenomics from irrigation water samples*

## *Methods*

*Virus strains and irrigation water samples*

Stool samples positive for norovirus GI, norovirus GII and astroviruses (HAstV, courtesy of Dr. Buesa, University of Valencia, Spain) and hepatitis E virus (HEV, courtesy of Dr. Alcaraz, Hospital Clínico Universitario, Valencia, Spain) were diluted to 10% wt/vol, in phosphate-buffered saline (PBS) containing 2M $NaNO_3$ (Fisher Scientific), 1% beef extract (Conda), 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. The cytopathogenic HM-175 strain (ATCC VR-140) of HAV, the human rotavirus (RV) strain Wa (ATCC VR-2018) and the MC0 strain (CECT 100000) of Mengovirus (MgV) were propagated in FRhK-4, MA-104 and HeLa cell monolayers, respectively. When cytopathic effects (CPE) were complete, cultures were frozen (−80°C) and then thawed three times, and cell debris removed by centrifugation at 3000 × g for 20 min. The supernatant, which contained viral particles, was stored at −80°C in aliquots. Seven archived irrigation water samples, originally collected from Balsicas (Murcia, Spain), positive for norovirus were used for the validation of the optimized procedure.

*Sample processing and nucleic acid extraction*

Initially, a mock sample (referred as mock-PBS) was prepared by artificially inoculating 6.0 log international units (IU)/ml norovirus GI, 5.7 log IU/ml norovirus GII, 5.3 log IU/ml HAV, 6.5 log PCR units (PCRU)/ml RV, 6.7 log PCRU/ml MgV, 5.8 log PCRU/ml HAstrV and, 5.2 log IU/ml HEV in PBS. An irrigation water sample positive for norovirus GI (3.83 log IU/ml) and GII (3.92 log IU/ml) was also included. Both samples were processed using five different protocols for the enrichment of viral reads (**Figure 1**). Firstly, the NetoVIR procedure (P0) was evaluated as described by Conceição-Neto et al. (2015) with some modifications. Briefly, 500 µl of the sample was homogenized with MP Fast Prep24 5G equipment (MP Biomedicals) for 40 s at speed 6.0 (as according to “Waste Water” program procedure available into the device) and centrifuged at 16,000 for 3 min. Then, 200 µl of supernatant was filtered with 0.8 µm PES filters (Sartorius). After filtration, samples were treated with benzonase and micrococcal nuclease. Total DNA/RNA was extracted using the NucleoSpin®RNA virus kit (Macherey-Nagel GmbH & Co.) without carrier RNA.

Then, four different enzymatic treatments were evaluated. In particular, 1U Turbo DNase (Invitrogen), 7U RNase A (Qiagen), 52.54U benzonase (Millipore) and 2,000U micrococcal nuclease (New England Biolabs) were applied to the samples as detailed herein. Procedure number 1 (P1) consisted of applying all the enzymes at once for 2 hours at 37°C, and, final inactivation with 0.5 nM EDTA. Procedure number 3 (P3) consisted of treating the samples with 52.54U benzonase and 2,000U nuclease for 2 hours at 37°C, then 14 µl of TURBO DNase buffer, 1U of Turbo DNase and 7U of RNase A were added and incubated for 1 hour at 37°C. To stop enzymatic activities, 28 µl of DNase Inactivation Reagent (Invitrogen) were added to each sample. Procedures number 2 (P2) and number 4 (P4) included the enzymatic treatments as described for P1 and P3, respectively, followed by a further concentration step with the Amicon® Ultra-0.5 filters (Merck KGaA). Finally, total RNA and DNA were extracted concomitantly using the NucleoSpin®RNA virus kit (Macherey-Nagel) without carrier RNA.

For comparative purposes, viral mock-PBS and irrigation water samples were directly extracted using the NucleoSpin RNA virus kit (Macherey-Nagel) following manufacturer’s instructions, including the use of the RNA carrier, and quantified as described below.

*RT-qPCR quantification*

Primers, probes and RT-qPCR conditions for norovirus GI, norovirus GII, RV, HAV, HEV, mengovirus and, HAstV quantification have previously been reported (Randazzo et al., 2019). Reverse transcription–quantitative polymerase chain reaction (RT-qPCR) assays were carried out in the LightCycler® 480 System (Roche, Switzerland) and analyzed by LightCycler® 480 software v1.5.0 (Roche). Standard curves were determined by the use of the Public Health England (PHE) reference materials for microbiology for norovirus GI (batch number 0122-17), norovirus GII (batch number 0247-17) and HAV (batch number 0261-2017), and by the first WHO international standard for HEV nucleic acid amplification technique (NAT)-based assays for HEV (code 6329/10; Paul-Ehrlich-Institut). HAstrV, RV and MgV standard curves were generated by amplifying 10-fold serial end-point dilutions of stool (HAstrV and RV) and cell-culture (MgV) suspensions in quintuplicates and calculating the numbers of PCRU/ml.

*Library preparation and sequencing*

Three different library preparation protocols for random sequencing were also compared (**Figure 1**). Two libraries were generated from 1 to 50 ng of DNA-RNA sample using the ScriptSeq v2 RNA-Seq Library Preparation Kit technology (Illumina) with slight modifications. In detail, the reverse transcriptase enzyme from the kit was substituted by the Reverse Transcriptase AMV of Thermo Scientific, an initial 5-min denaturation step at 95°C was included, and the number of PCR cycles was increased up to 18 (L18) or 20 (L20) (instead of 10–15 cycles as suggested by manufacturer) to boost library concentration. A third library (WTA) was prepared by Complete Whole Transcriptome Amplification kit (Sigma-Aldrich) combined with Nextera XT library preparation kit (Illumina), as described in the NetoVIR protocol (Conceição-Neto et al., 2015). Libraries were normalized, pooled and sequenced using the NextSeq™ 500 system (Illumina), following the manufacturer's protocol, 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 (Martin, 2011) with a minimum overlap length between read and adapter of five nucleotides and a maximum error rate allowed 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. Cleaned paired-end reads were merged in single reads using FLASH v1.2.11 (Magoč and Salzberg, 2011) allowing outies. Merged sequences were taxonomically annotated with BLASTn with an e-value of 1e-5 using the 'Reference Viruses Representative Genomes' database (henceforth 'RefGenomes' database) from the National Center for Biotechnology Information (NCBI) (ftp://ftp.ncbi.nih.gov/blast/db/ref_viruses_rep_genomes.tar.gz, released on February 08, 2019). Richness values were calculated using the improved non-parametric Chao index (Chao and Chiu, 2016) with SpadeR software (https://chao.shinyapps.io/SpadeR/). Only taxonomic affiliations that covered more than 50% of the query sequence and the percentage of similarity higher than 70% were used for virome analyses.

For comparative purposes, samples were also taxonomically annotated by using two additional tools: (i) an in-house database constructed with the depurated nucleotide sequences together with the viral genomes available at NCBI database (https://www.ncbi.nlm.nih.gov/nuccore/?term=viruses%5Borganism%5D) applying a BLASTn analysis, and (ii) the ViromeScan software (Rampelli et al., 2016) exploiting default parameters (Figure 1).

**Results** – Objective 1

We initially focused on the enrichment of viral particles in irrigation water and mock-PBS samples applying the NetoVir protocol (Conceição-Neto et al., 2015). Performance of the procedure was determined by RT-qPCR and metagenomics sequencing. Additionally, we tested and compared different methodologies of sample preprocessing, sequencing library construction and data analysis (**Figure 1**). A total of five preprocessing methods and three sequencing libraries were tested on the mock-PBS and irrigation water samples. The percentage of annotated viral sequences from the total reads obtained after quality trimming and read merging with the RefGenomes Viral database are shown in **Figure 2**. The percentages of viral reads for the viral mock-PBS sample ranged from 0.47% in the P0-L20 sample to 3.08% in P3-L20. The highest percentages of viral reads were obtained with libraries of 20 cycles in the samples where the four enzymes were added in two steps (3.08% in P3-L20 and 1.51% in P4-L20), being higher when filtering with the Amicon® Ultra-0.5 filters was not performed (P3). The percentages of viral reads identified by the RefGenomes database in irrigation water samples ranged from 0.02 to 0.28%, for P4-WTA and P2-WTA, respectively. The highest percentages of viral reads were obtained in samples P1-L20 (0.08%), P2-L20 (0.11%) and P2-WTA (0.28%). However, the high percentages found in these samples corresponded mostly to reads identified as rotavirus in P1-L20 (55.38%) and P2-L20 (60.37%), and to Escherichia phages Lambda (29.40%) and M13 (35.41%) in P2-WTA.

To further evaluate the selected procedures for irrigation water samples, seven irrigation water samples were analyzed using the following protocols: extraction procedures P0, P3 and P4 with the ScriptSeq v2 RNA-Seq Library Preparation Kit at 18 PCR cycles and the procedure P0 with the same library kit at 20 PCR cycles. Additionally, the concentration of enteric viruses present in each sample was quantified by RT-qPCR. After the analysis of the results obtained, the work-flow P0-L20 (the NetoVir preprocessing which includes a filtration and nuclease digestion procedure prior to DNA/RNA extraction followed by the preparation of sequencing libraries for each individual sample) combined with an in-house database was selected because of the higher percentage of identified sequences and the better detection of relevant human enteric viruses (**Figure 3**).

**Objectives 2 & 3:** *Determination of the viral community composition of samples previously analyzed that tested positive or negative for the presence of human pathogenic viruses & Identification of specific viral species or groups whose presence/abundance correlates with the occurrence of human pathogenic viruses in stools, irrigation waters and produce.*

***Methods***

*Newly collected irrigation water and produce samples*

Two-hundred ml samples of irrigation water were collected in Valencia (Spain) and each sample was inoculated with mengovirus vMC0 (CECT 100000), used as a process control. Water samples were processed using the aluminum-based precipitation protocol described elsewhere (Randazzo et al., 2019). Water concentrates were resuspended in 1 ml of PBS (pH 7.4) and stored at −80°C. Lettuce samples were collected and processed using the virus concentration procedure described in the ISO15216-1:2017 standard; mengovirus was used as a process control.

*Sample processing and analysis*

Viral particles from archived and newly collected samples were enriched from water and produce concentrates following the NetoVIR protocol. Capsid protected viral nucleic acids were

extracted with the NucleoSpin®RNA virus kit (Macherey-Nagel GmbH & Co.), according to the manufacturer's instructions, without adding carrier RNA. Nucleic acids were eluted in 100 µl of RNase-free water. Libraries were generated from 1 to 50 ng of a DNA-RNA sample using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, USA). Libraries were normalized, pooled, and sequenced using the NextSeq™ 500 platform (Illumina), following the manufacturer's protocol, with a configuration of 150 cycles paired-end reads. Sequencing was performed by Lifesequencing S.L. (Valencia, Spain). Library preparation and sequencing and data analysis was performed as described above.

*Virus quantification*

For virus quantification an optimized viability RT-qPCR was applied as previously described (Randazzo et al., 2019). In brief, 150-µl water concentrates were added to 50 µM PMAxx (Biotium, USA) and 0.5% Triton 100-X (Thermo Fisher Scientific) and incubated in the dark at room temperature (RT) for 10 min at 150 rpm. Then, samples were exposed to photo-activation using a photo-activation system (Led-Active Blue, GenIUL) for 15 min. RNA was extracted using the NucleoSpin® RNA virus kit according to the manufacturer's instructions, including the Plant RNA Isolation Aid (Ambion) pretreatment (Randazzo et al., 2018, 2016). Primers, probes and RT-qPCR conditions for norovirus GI, norovirus GII, RV, HAV, HEV, mengovirus and HAstV quantification have been previously reported (Randazzo et al., 2019).

Quantification of crAssphage was performed using the primer set CPQ_064 described by Stachler et al. (2017). PCR conditions were: a first denaturation step of 30 seconds at 95°C followed by 45 cycles of 5s at 95°C and 30s at 60°C. The Premix Ex Taq master mix for probe-based real-time PCR kit (Takara) was used for the reaction. For the crAssphage quantification, the standard curve was performed with a customized gBlock® fragment (Integrated DNA Technologies, Spain) of 228 bp that contained the crAssphage sequences used for amplification.

*Correlation analyses*

Correlations between results obtained by RT-qPCR quantifications of noroviruses GI and GII, RV, HAstrV, and HAV, and qPCR quantifications of crAssphage along with the number of reads obtained for these viruses and viruses proposed as fecal contamination markers were calculated using the R package Hmisc v4.2-0 (https://CRAN.R-project.org/package=Hmisc) following the Pearson method. Significance was set at 0.05. Representation of correlation matrix values was performed with the R library corrplot v0.84 (https://CRAN.R-project.org/package=corrplot).

**Results** – Objectives 2 & 3

Initially, quantification of the most relevant human enteric viruses was performed by viability RT-qPCR. In parallel, viral particles were enriched from water and produce concentrates using the protocol defined in Objective 1. Viromes from irrigation water and produce were characterized with high abundance of bacteriophages from the *Microviridae* and *Siphoviridae* families and plant viruses from the *Virgaviridae* family (**Figure 5**). As the the ScriptSeq v2 RNA-Seq Library Preparation Kit was discontinued by the provider during the timeframe of the project, several alternative kits were evaluated (**Figure 4**). According to percentage of viral reads and virus diversity, the NEBNext kit was selected for further processing of the samples.

Overall, 254 virome profiles were obtained from stool, produce and irrigation water samples (**Figure 5**). Samples were also analyzed for the presence of human pathogenic viruses, i.e., norovirus GI, GII, HAV, RV and HastV. In irrigation water samples, a total of 35%, 24%, 4%,

48%, 29% of the samples were positive for norovirus GI, GII, HAV, RV and HastV, respectively. In produce samples, only HAV and RV were detected with 2% and 5% of positive samples, respectively. Finally, the percentage of positive samples for norovirus GI, GII, RV and HastV in stool samples were 11%, 6%, 39% and 11%, respectively.

We then focused on searching proposed fecal viral indicators, such as crAssphage, pepper mild mottle virus (PMMoV), picobirnavirus, torque teno virus, human polyomavirus, human adenovirus, enterovirus, and somatic and F-specific coliphages (families *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Microviridae*, *Leviviridae* and *Inoviridae*). Initially, samples were classified as positive or negative for human norovirus, and the most common viral fecal indicators were searched within the metagenomics data of each individual sample (**Table 1**). The most common viral fecal indicators were *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Microviridae*, present in all norovirus-positive samples, however they were also highly abundant in the norovirus-negative samples. Similar results were obtained for PMMoV and human picobirnaviruses. According to these results, all the above-mentioned viruses are not suitable indicators for assessing norovirus contamination, as a viral indicator should be not present in norovirus-negative samples and present in the norovirus-positive samples. Following this criteria, crAssphage reads were present in 89% of the norovirus-positive samples and in 33% of the norovirus-negative samples. Moreover, crAssphage reads were present in 82% of the enteric virus-positive samples while crAssphage reads were only in 23 % of the enteric virus-negative samples (**Table 1**). When metagenomics data was compared in positive and negative sub-set samples, other potential indicators were identified. Among them, EBPR podovirus 2, Aeromonas phage4_L372X, Pseudomonas phage NP1, EBPR siphovirus 1, Bordetella phage vB_BbrM_PHB04, and Aeromonas phage 62AhydR11PP were highly abundant in positive samples (from 65 to 89%) while in low abundance in negative samples (from 8 to 29%).

As indicators should be readily detected, a previously described qPCR procedure for crAssphage detection (Stachler et al., 2017) was performed in all samples, and presence and absence was correlated with the RT-qPCR results of human enteric viruses (**Figure 6**). Overall, crAssphage qPCR results show positive correlation, with values from 0.51 and 0.63, with RT-qPCR of norovirus GI, GII RV and HastV. According to qPCR data, crAssphage was present in 93% of the norovirus-positive samples and only in 23 % of the norovirus-negative samples.

Interestingly, as RNA viruses infecting plants as natural hosts (i.e., *Virgaviridae*) have been detected in irrigation water by the methods developed in the frame of this project, side information on their occurrence could be obtained and used to limit their environmental transmission.

## Outcomes and Accomplishments

- A total of 138 irrigation water and produce samples were collected from two different fresh produce growing regions in Spain over a two-year period.
- A detailed procedure for the concentration of human pathogenic viruses in irrigation water, produce and stool samples was established.
- Occurrence of human enteric viruses (i.e., norovirus GI, GII, RV, HAV and HAstV) and crAssphage in 192 irrigation water and produce samples was determined.
- Suitable methods (pre-processing, kit for library preparation and bioinformatics pipeline) for the characterization of the virome in irrigation water, produce and stool samples were assessed and selected based on 45 combinations for a mock-PBS sample and 7 irrigation water samples.
- Viral community composition (virome) from archived and newly collected samples (a total of 254) was characterized.
- Correlation between occurrence of human enteric viruses and presence of proposed viral indicators was determined.
- Identification of potential new viral indicators from metagenomics data: EBPR podovirus 2, Aeromonas phage4_L372X, Pseudomonas phage NP1, EBPR siphovirus 1, Bordetella phage vB_BbrM_PHB04, and Aeromonas phage 62AhydR11PP.
- Correlation of crAssphage and human enteric virus occurrence was established.

## Summary of Findings and Recommendations

Detection of human enteric viruses would be the ideal option for determining the virological quality of irrigation waters and produce; but at present, this is neither practical nor feasible in routine testing for most of the laboratories. The rapid and extensive development of next-generation sequencing (NGS) has opened up more opportunities to advance understanding in virome characterization. Virome characterization via metagenomics comprises the general steps of sampling, sample processing, sequencing, and data analysis. In this study, the selected metagenomics workflow was already proven suitable for the virome characterization of irrigation water, produce and stool samples. These samples were also analyzed for the presence of the most relevant human enteric viruses (i.e., norovirus GI, GII, RV, HAV and HAstV) by viability RT-qPCR, and their presence was correlated with metagenomics data in order to select a suitable virus indicator. Overall, our data indicated that the recently proposed viral indicator, crAssphage, significantly correlated with the presence of human enteric viruses. To confirm this point, crAssphage occurrence was also determined in the same samples, confirming these results, although 100% correlation was not achieved. Given that an ideal indicator should be easily detectable by rapid and inexpensive laboratory tests and inform ideally on viral potential infectivity, the detection of crAssphage by molecular methods lacks in matching these requirements, especially because it does not infer infectivity.

Metagenomics data from this study also suggested that EBPR podovirus 2, Aeromonas phage4_L372X, Pseudomonas phage NP1, EBPR siphovirus 1, Bordetella phage vB_BbrM_PHB04, and Aeromonas phage 62AhydR11PP are alternative suitable indicators for contamination by enteric viruses. However further research is needed in order to develop rapid methods for their detection and assess their correlation using other methods as qPCR or culture.

## APPENDICES

### Publications (in preparation)

A. Pérez-Cataluña, W. Randazzo, J. F. Martínez-Blanch, F. M. Codoñer, D. Carrillo, G. Sánchez. Optimized procedure for virome analysis in irrigation water samples. *In preparation.*

### Presentations

G. Sánchez: Preliminary results from this project have been presented by the Principal Investigator at the annual CPS Research Symposium in 2018 (lightning round talk and poster) and 2019 (full presentation). This work has also been presented in a video on the CPS webpage. A presentation of the final results will be given at the 2020 CPS Research Symposium in La Jolla, CA.

A. Pérez-Cataluña, W. Randazzo, F. M. Codoñer, J. F. Martínez, D. Carrillo, G. Sánchez. 2018. Comparación de diferentes métodos de tratamiento y análisis metagenómico de muestras de agua para detección de virus. VII Reunión del Grupo de Taxonomía, Filogenia y Biodiversidad y XII del Grupo de Microbiología del Medio Acuático, Sitges (Spain).

A. Pérez-Cataluña, W. Randazzo, G. Sánchez. 2019. Viral metagenomics in irrigation waters: assemblers impact. 3rd Annual Meeting of the European Virus Bioinformatics Center Glasgow (UK).

### Budget Summary

Total funds awarded to the project were $301,647, and all funds are expected to be used.

**Table and Figures** (see below)

**Table 1 and Figures 1–6**

Table 1. Correlation table between viral reads obtained by metagenomics and presence or absence of norovirus and human enteric viruses (i.e., norovirus, RV, HAstV and HAV by RT-qPCR).

| | Norovirus | | | Human enteric viruses (norovirus, RV, HAstV or HAV) | |
|---|---|---|---|---|---|
| **Proposed indicators** | Positive[1] | Negative[2] | | Positive[3] | Negative[4] |
| *Myoviridae* | 100% | 94% | | 100% | 92% |
| *Siphoviridae* | 100% | 97% | | 100% | 96% |
| *Podoviridae* | 100% | 89% | | 100% | 86% |
| *Microviridae* | 100% | 93% | | 99% | 91% |
| *Leviviridae* | 41% | 36% | | 48% | 26% |
| Inoviridae | 88% | 68% | | 85% | 66% |
| crAssphage | 89% | 33% | | 82% | 23% |
| Pepper mild mottle virus | 80% | 86% | | 85% | 83% |
| Torque Teno virus | 0% | 1% | | 0% | 1% |
| Human picobirnavirus | 96% | 93% | | 97% | 90% |
| Human polyomavirus 1 | 1% | 0% | | 1% | 0% |
| Human adenovirus | 58% | 78% | | 71% | 77% |
| Human enterovirus | 2% | 0% | | 1% | 0% |
| **Potential new indicators** | | | | | |
| EBPR podovirus 2 | 93% | 40% | | 89% | 28% |
| Aeromonas phage4_L372X | 85% | 38% | | 80% | 29% |
| Pseudomonas phage NP1 | 76% | 18% | | 69% | 9% |
| EBPR siphovirus 1 | 76% | 21% | | 71% | 10% |
| Bordetella phage vB_BbrM_PHB04 | 75% | 22% | | 65% | 17% |
| Aeromonas phage 62AhydR11PP | 74% | 18% | | 68% | 8% |

[1] Percentage of samples containing reads of the indicator in samples that tested positive for norovirus by RT-qPCR.
[2] Percentage of samples containing reads of the indicator in samples that tested negative for norovirus by RT-qPCR.
[3] Percentage of samples containing reads of the indicator in samples that tested positive for norovirus, RV, HAstV or HAV by RT-qPCR.
[4] Percentage of samples containing reads of the indicator in samples that tested negative for norovirus, RV, HAstV or HAV by RT-qPCR.

Figure 1. Schematic representation of the different methods used in preprocessing, library construction and data analysis compared in this project.

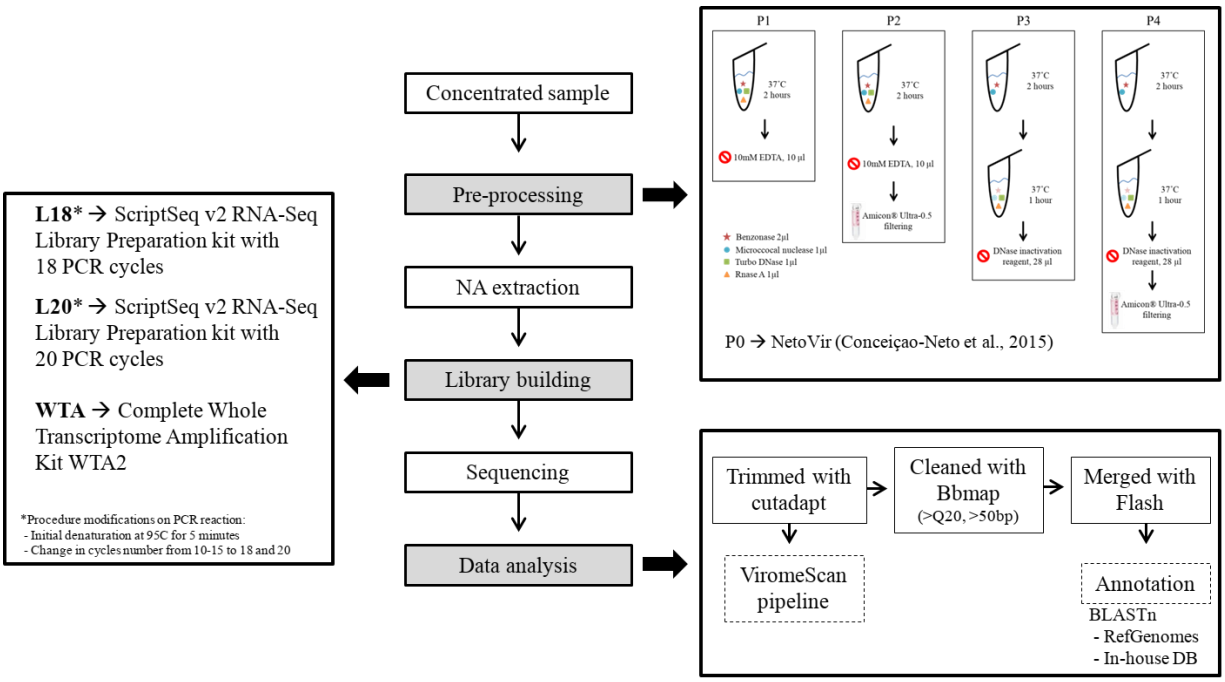


Figure 2. Heat-map showing the viral diversity on viral mock-PBS (A) and irrigation water (B) samples.

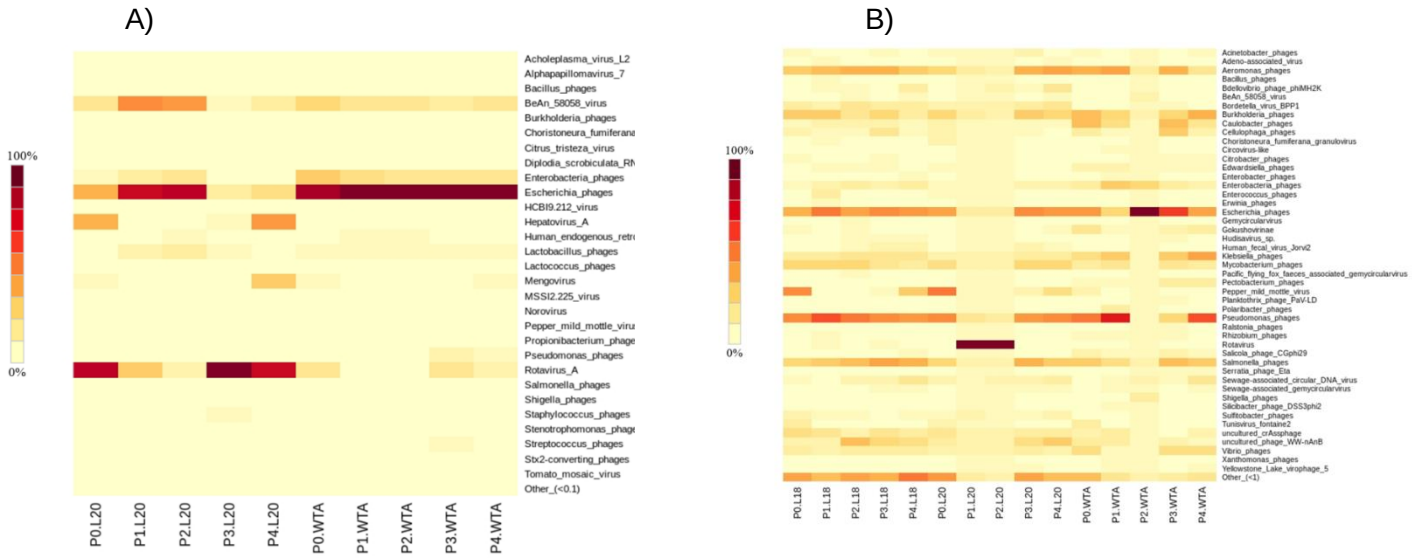



Figure 3. Percentage of viral reads in irrigation water samples used for the method validation. Asterisks indicate Chao richness.

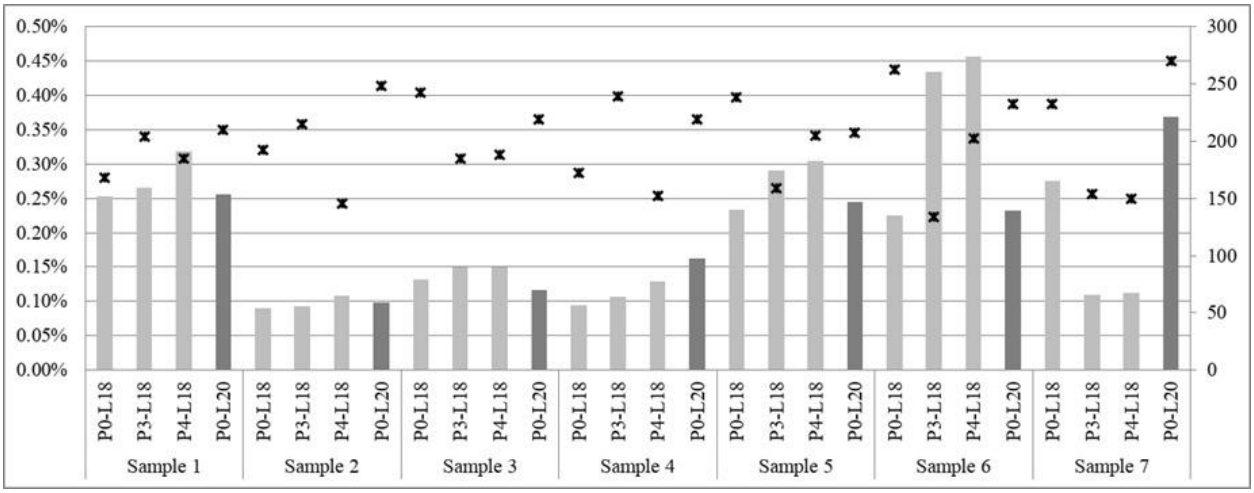
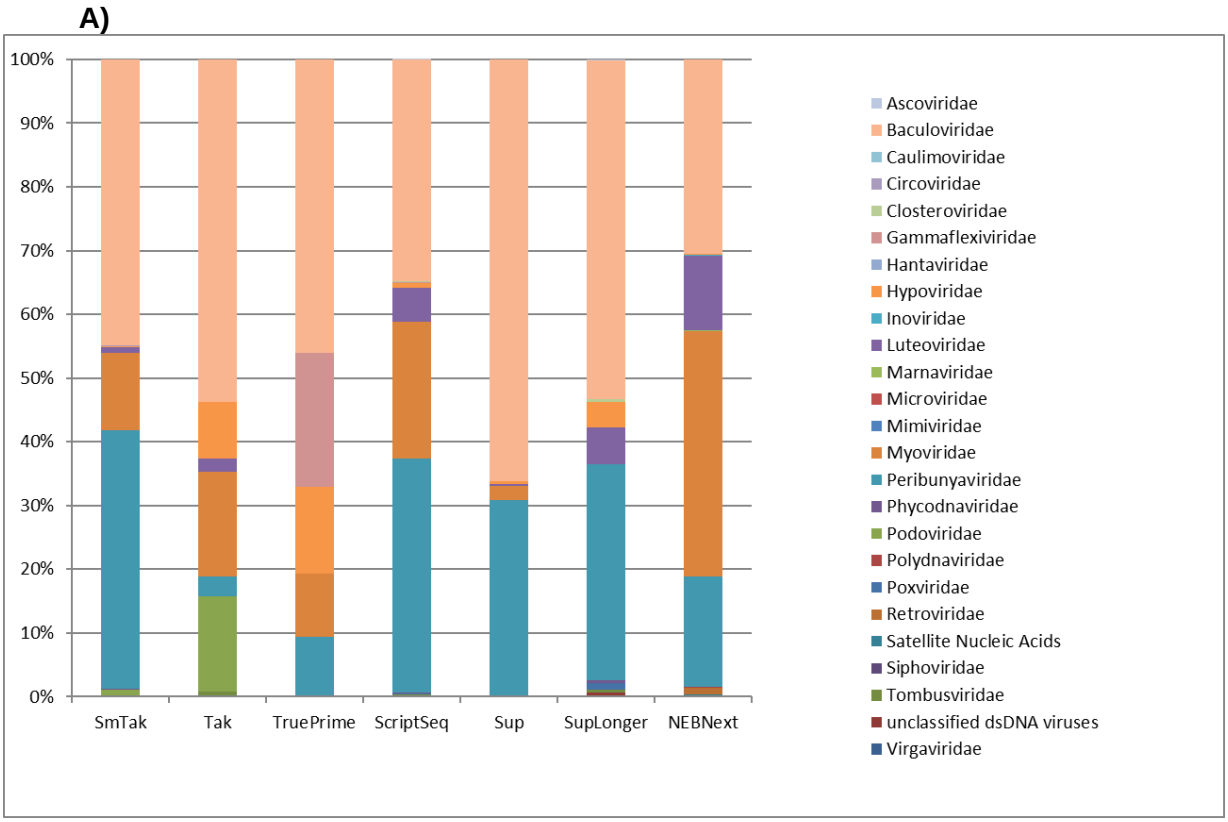


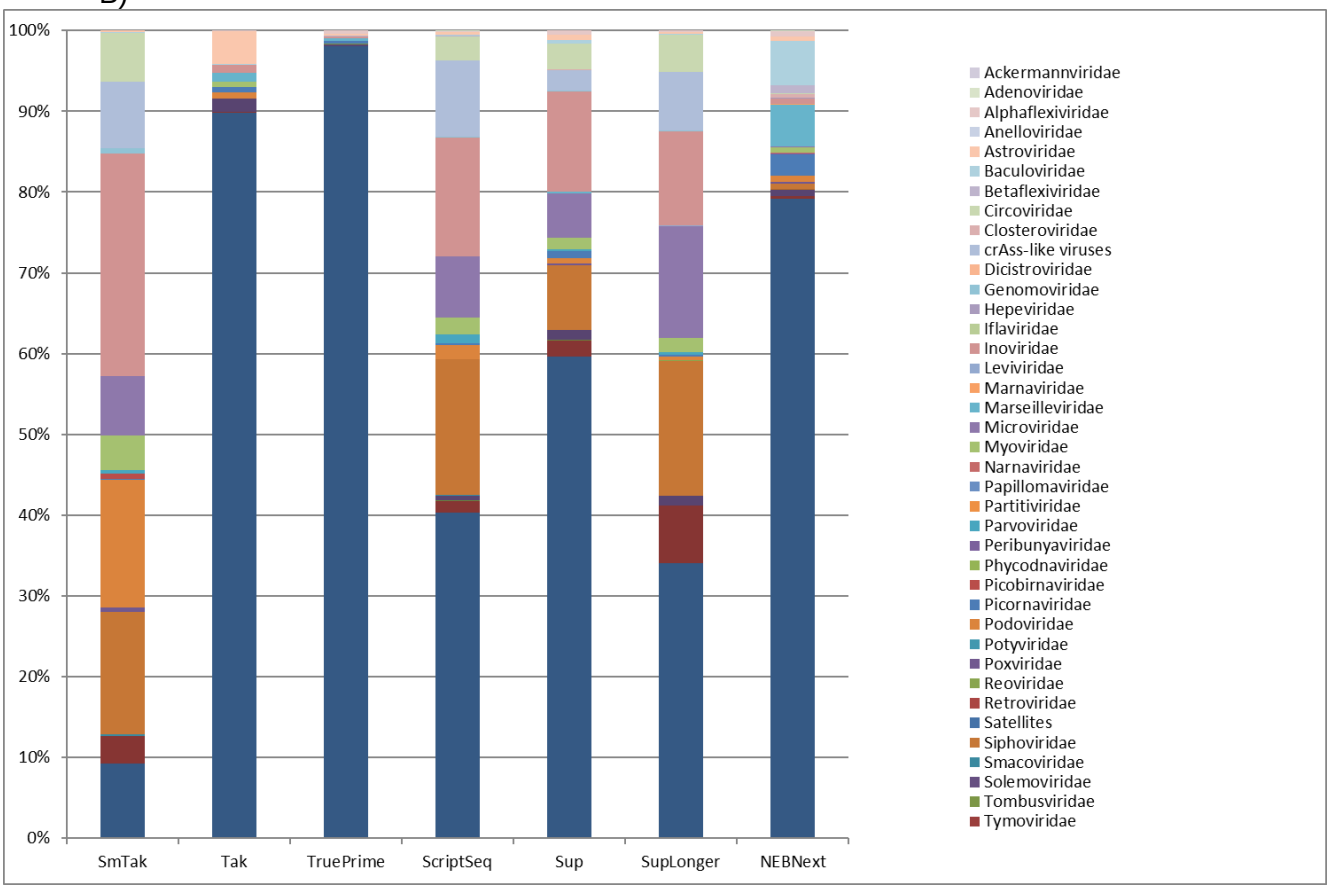
Figure 4. Viral diversity using different kits for library preparation: A) zucchini, B) irrigation water




SmTAK: SMART-Seq® Stranded Kit with Superscript IV and Takara enzimes; Tak: SMART-Seq® Stranded Kit with Takara enzime; TruePrime: SMART-Seq® Stranded Kit with Sygnis enzyme; ScriptSeq: ScriptSeq v2 RNA-Seq Library Preparation Kit; Sup: SMART-Seq® Stranded Kit with Superscript IV enzyme; SupLonger: as 'Sup' but selecting long fragments for sequencing; NEBNext: NEBNext Ultra II DNA Library Prep Kit.

Figure 4 cont'd. Viral diversity using different kits for library preparation: A) zucchini, B) irrigation water

B)




SmTAK: SMART-Seq® Stranded Kit with Superscript IV and Takara enzimes; Tak: SMART-Seq® Stranded Kit with Takara enzime; TruePrime: SMART-Seq® Stranded Kit with Sygnis enzyme; ScriptSeq: ScriptSeq v2 RNA-Seq Library Preparation Kit; Sup: SMART-Seq® Stranded Kit with Superscript IV enzyme; SupLonger: as 'Sup' but selecting long fragments for sequencing; NEBNext: NEBNext Ultra II DNA Library Prep Kit.

Figure 5. Heat-map showing the viral diversity of taxons with percentages higher than 1% from norovirus positive (A) and negative samples (B).

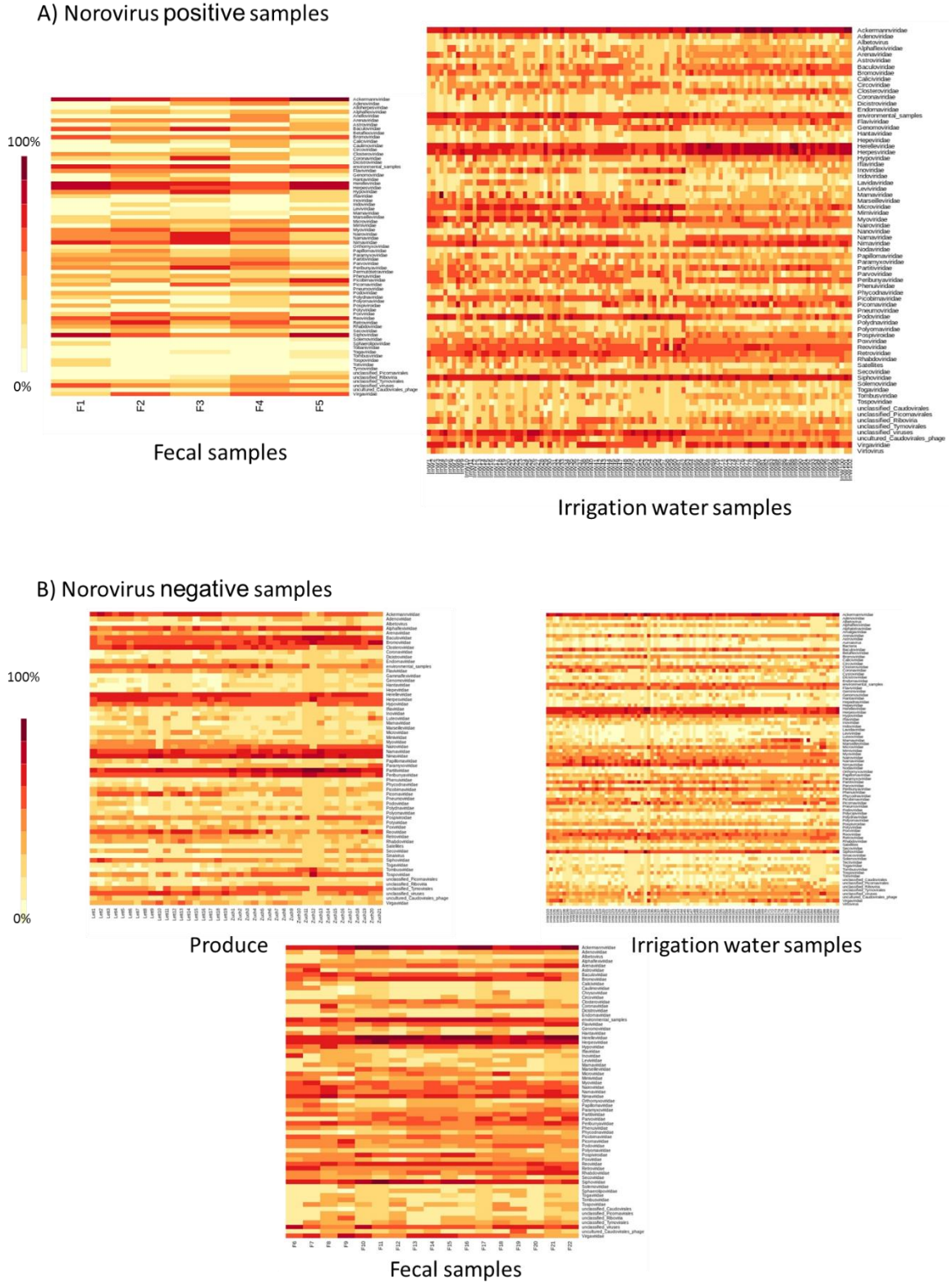
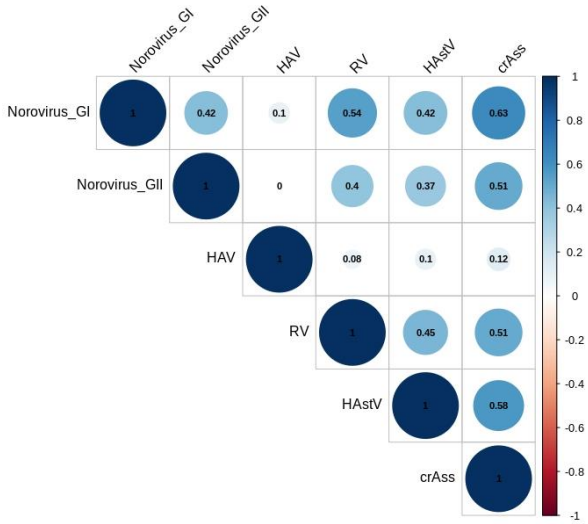
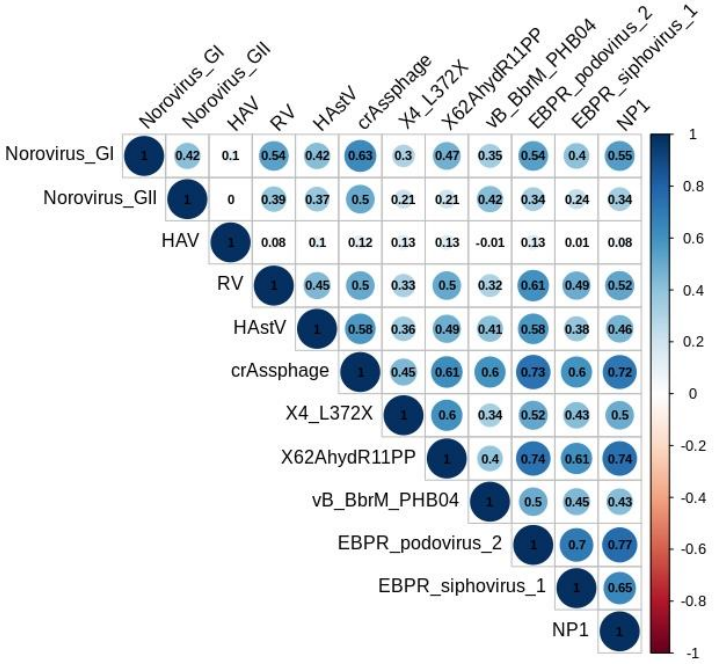


Figure 6. Correlation matrix between (A) crAssphage quantification obtained by qPCR and quantifications of viral pathogens by RT-qPCR and, (B) quantification of crAssphage and human enteric viruses obtained by qPCR/RT-qPCR and viral reads from potential new indicators obtained by metagenomics.

A)




B)



## References Cited

Ashbolt, N.J., 2015. Microbial Contamination of Drinking Water and Human Health from Community Water Systems. Curr. Environ. Heal. reports. https://doi.org/10.1007/s40572-014-0037-5

Baert, L., Mattison, K., Loisy-Hamon, F., Harlow, J., Martyres, A., Lebeau, B., Stals, A., Van Coillie, E., Herman, L., Uyttendaele, M., 2011. Review: norovirus prevalence in Belgian, Canadian and French fresh produce: a threat to human health? Int. J. Food Microbiol. 151, 261–9. https://doi.org/10.1016/j.ijfoodmicro.2011.09.013

Bennett, S.D., Sodha, S. V., Ayers, T.L., Lynch, M.F., Gould, L.H., Tauxe, R. V., 2018. Produce-associated foodborne disease outbreaks, USA, 1998-2013. Epidemiol. Infect. 146, 1397–1406. https://doi.org/10.1017/S0950268818001620

Chao, A., Chiu, C.-H., 2016. Nonparametric Estimation and Comparison of Species Richness, in: ELS. John Wiley & Sons, Ltd, Chichester, UK, pp. 1–11. https://doi.org/10.1002/9780470015902.a0026329

Chandrasekaran, S., Jiang, S. C., 2018. A dynamic transport model for quantification of norovirus internalization in lettuce from irrigation water and associated health risk. Sci. Total Environ. 643, 751–761. doi:10.1016/j.scitotenv.2018.06.158

Cheong, S., Lee, C., Song, S.W., Choi, W.C., Lee, C.H., Kim, S.J., 2009. Enteric viruses in raw vegetables and groundwater used for irrigation in South Korea. Appl. Environ. Microbiol. 75, 7745–7751. https://doi.org/10.1128/AEM.01629-09

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C.K., Lavigne, R., Maes, P., Van Ranst, M., Heylen, E., Matthijnssens, J., 2015. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Sci. Rep. 5, 16532. https://doi.org/10.1038/srep16532

Dentinger, C.M., Bower, W.A., Nainan, O.V., Cotter, S.M., Myers, G., Dubusky, L.M., Fowler, S., Salehi, E.D.P., Bell, B.P., 2001. An Outbreak of Hepatitis A Associated with Green Onions. J. Infect. Dis. 183, 1273–1276. https://doi.org/10.1086/319688

Dicaprio, E., Ma, Y., Purgianto, A., Hughes, J., Li, J., 2012. Internalization and dissemination of human norovirus and animal caliciviruses in hydroponically grown romaine lettuce. Appl. Environ. Microbiol. 78(17), 6143–6152. https://doi.org/10.1128/AEM.01081-12

El-Senousy, W.M., Costafreda, M.I., Pintó, R.M., Bosch, A., 2013. Method validation for norovirus detection in naturally contaminated irrigation water and fresh produce. Int. J. Food Microbiol. 167, 74–79. https://doi.org/10.1016/j.ijfoodmicro.2013.06.023

Ethelberg, S., Lisby, M., Böttiger, B., Schultz, A.C., Villif, A., Jensen, T., Olsen, K.E., Scheutz, F., Kjelsø, C., Müller, L., 2010. Outbreaks of gastroenteritis linked to lettuce, Denmark, January 2010. Eurosurveillance 15, 1. https://doi.org/10.2807/ese.15.06.19484-en

Falkenhorst, G., Krusell, L., Lisby, M., Madsen, S.B., Böttiger, B., Mølbak, K., 2005. Imported frozen raspberries cause a series of norovirus outbreaks in Denmark, 2005. Euro Surveill. 10. https://doi.org/10.2807/esw.10.38.02795-en

Felix-Valenzuela, L., Resendiz-Sandoval, M., Burgara-Estrella, A., Hernández, J., Mata-Haro, V., 2012. Quantitative Detection of Hepatitis A, Rotavirus and Genogroup I Norovirus by

RT-qPCR in Fresh Produce from Packinghouse Facilities. J. Food Saf. 32, 467–473. https://doi.org/10.1111/jfs.12009

Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R., Rose, J.B., 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl. Environ. Microbiol. 71, 3163–3170. https://doi.org/10.1128/AEM.71.6.3163-3170.2005

Hernández, F., Monge, R., Jiménez, C., Taylor, L., 1997. Rotavirus and hepatitis A virus in market lettuce (*Latuca sativa*) in Costa Rica. Int. J. Food Microbiol. 37, 221–223. https://doi.org/10.1016/S0168-1605(97)00058-5

Jofre, J., Lucena, F., Blanch, A.R., Muniesa, M., 2016. Coliphages as model organisms in the characterization and management of water resources. Water (Switzerland) 8, 1–21. https://doi.org/10.3390/w8050199

Kokkinos, P., Kozyra, I., Lazic, S., Bouwknegt, M., Rutjes, S., Willems, K., Moloney, R., de Roda Husman, A.M., Kaupke, A., Legaki, E., D'Agostino, M., Cook, N., Rzeżutka, A., Petrovic, T., Vantarakis, A., 2012. Harmonised investigation of the occurrence of human enteric viruses in the leafy green vegetable supply chain in three European countries. Food Environ. Virol. 4, 179–91. https://doi.org/10.1007/s12560-012-9087-8

Li, M., Baker, C.A., Danyluk, M.D., Belanger, P., Boelaert, F., Cressey, P., Gheorghe, M., Polkinghorne, B., Toyofuku, H., Havelaar, A.H., 2018. Identification of Biological Hazards in Produce Consumed in Industrialized Countries: A Review. J. Food Prot. 81, 1171–1186. https://doi.org/10.4315/0362-028X.JFP-17-465

López-Gálvez, F., Truchado, P., Sánchez, G., Aznar, R., Gil, M.I., Allende, A., 2016. Occurrence of enteric viruses in reclaimed and surface irrigation water: relationship with microbiological and physicochemical indicators. J. Appl. Microbiol. 121, 1180–1188. https://doi.org/10.1111/jam.13224

Machado-Moreira, B., Richards, K., Brennan, F., Abram, F., Burgess, C.M., 2019. Microbial Contamination of Fresh Produce: What, Where, and How? Compr. Rev. Food Sci. Food Saf. https://doi.org/10.1111/1541-4337.12487

Magoč, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–63. https://doi.org/10.1093/bioinformatics/btr507

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10. https://doi.org/10.14806/ej.17.1.200

McQuaig, S., Griffith, J., Harwood, V.J., 2012. Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. Appl. Environ. Microbiol. 78, 6423–6432. https://doi.org/10.1128/AEM.00024-12

Monge, R., Arias, M.L., 1996. [Presence of various pathogenic microorganisms in fresh vegetables in Costa Rica]. Arch. Latinoam. Nutr. 46, 292–294.

Pérez-Sautu, U., Sano, D., Guix, S., Kasimir, G., Pintó, R.M., Bosch, A., 2012. Human norovirus occurrence and diversity in the Llobregat river catchment, Spain. Environ. Microbiol. 14, 494–502. https://doi.org/10.1111/j.1462-2920.2011.02642.x

Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., Candela, M., 2016. ViromeScan: A new tool for metagenomic viral community profiling. BMC Genomics 17, 1–9. https://doi.org/10.1186/s12864-016-2446-3

Randazzo, W., López-gálvez, F., Allende, A., Aznar, R., Sánchez, G., 2016. International Journal of Food Microbiology Evaluation of viability PCR performance for assessing norovirus infectivity in fresh-cut vegetables and irrigation water. Int. J. Food Microbiol. 229, 1–6. https://doi.org/10.1016/j.ijfoodmicro.2016.04.010

Randazzo, W., Vásquez-García, A., Bracho, M.A., Alcaraz, M.J., Aznar, R., Sánchez, G., 2018. Hepatitis E virus in lettuce and water samples: A method-comparison study. Int. J. Food Microbiol. 277, 34–40. https://doi.org/10.1016/J.IJFOODMICRO.2018.04.008

Rosario, K., Symonds, E.M., Sinigalliano, C., Stewart, J., Breitbart, M., 2009. Pepper mild mottle virus as an indicator of fecal pollution. Appl. Environ. Microbiol. 75, 7261–7267. https://doi.org/10.1128/AEM.00410-09

Rusiñol, M., Fernandez-Cassi, X., Hundesa, A., Vieira, C., Kern, A., Eriksson, I., Ziros, P., Kay, D., Miagostovich, M., Vargha, M., Allard, A., Vantarakis, A., Wyn-Jones, P., Bofill-Mas, S., Girones, R., 2014. Application of human and animal viral microbial source tracking tools in fresh and marine waters from five different geographical areas. Water Res. 59, 119–129. https://doi.org/10.1016/j.watres.2014.04.013

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., Shanks, O.C., 2017. Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. Environ. Sci. Technol. 51, 9146–9154. https://doi.org/10.1021/acs.est.7b02703

Tian, P., Yang, D., Shan, L., Wang, D., Li, Q., Gorski, L., Lee, B.G., Quiñones, B., Cooley, M.B., 2017. Concurrent Detection of Human Norovirus and Bacterial Pathogens in Water Samples from an Agricultural Region in Central California Coast. Front. Microbiol. 8, 1560. https://doi.org/10.3389/fmicb.2017.01560

Torok, V.A., Hodgson, K.R., Jolley, J., Turnbull, A., McLeod, C., 2019. Estimating risk associated with human norovirus and hepatitis A virus in fresh Australian leafy greens and berries at retail. Int. J. Food Microbiol. 309, 108327. https://doi.org/10.1016/j.ijfoodmicro.2019.108327