Water Research 203 (2021) 117568

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity

Evelien M. Adriaenssens [a,b,*], Kata Farkas [c,d], James E. McDonald [c], David L. Jones [c,e], Heather E. Allison [a], Alan J. McCarthy [a]

[a] Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK
[b] Quadram Institute Bioscience, Norwich, NR4 7UQ, UK
[c] School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK
[d] School of Ocean Sciences, Bangor University, Bangor, LL59 5AB, UK
[e] UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6009, Australia

ARTICLE INFO

Keywords:
Viromics
viral diversity
wastewater viruses
aquatic viruses
shellfish viruses
wastewater contamination
virus ecology

ABSTRACT

The discharge of wastewater-derived viruses in aquatic environments impacts catchment-scale virome composition. To explore this, we used viromic analysis of RNA and DNA virus-like particles to holistically track virus communities entering and leaving wastewater treatment plants and the connecting river catchment system and estuary. We reconstructed >40 000 partial viral genomes into 10 149 species-level groups, dominated by dsDNA and (+)ssRNA bacteriophages (*Caudoviricetes* and *Leviviricetes*) and a small number of genomes that could pose a risk to human health. We found substantial viral diversity and geographically distinct virus communities associated with different wastewater treatment plants. River and estuarine water bodies harboured more diverse viral communities in downstream locations, influenced by tidal movement and proximity to wastewater treatment plants. Shellfish and beach sand were enriched in viral communities when compared with the surrounding water, acting as entrapment matrices for virus particles. Extensive phylogenetic analyses of environmental-derived and reference sequences showed the presence of human-associated sapovirus GII in all sample types, multiple rotavirus A strains in wastewater and a diverse set of picorna-like viruses associated with shellfish. We conclude that wastewater-derived viral genetic material is commonly deposited in the environment and can be traced throughout the freshwater-marine continuum of the river catchment, where it is influenced by local geography, weather events and tidal effects. Our data illustrate the utility of viromic analyses for wastewater- and environment-based ecology and epidemiology, and we present a conceptual model for the circulation of all types of viruses in a freshwater catchment.

1. Introduction

Viruses are the most abundant biological entities in terrestrial and aquatic biomes, but their origin, distribution and potential to spread disease via watercourses is poorly understood (Roux et al., 2020). Previous research has demonstrated that wastewater contains a plethora of viruses, including human-pathogenic and zoonotic viruses, and that wastewater treatment processes do not remove human viruses with sufficient efficacy (Da Silva et al., 2007; Farkas et al., 2018b; Fong et al., 2010; Girones et al., 2010; Gomes et al., 2019; Gulino et al., 2020; Hellmér et al., 2014; Kitajima et al., 2014; Prado et al., 2019; Qiu et al., 2015; Sidhu et al., 2017). Viral abundance, behaviour, infectivity and fate remain poorly understood because of knowledge gaps in the ecology and connectivity of viromes across human populations and the freshwater-marine continuum.

The current gold standard method for investigating enteric viruses in the environment is q(RT)-PCR, a technique that provides reliable quantitative information on the presence of the genomic material of target viruses, but requires prior knowledge on the identity of the virus and its genome sequence (Farkas et al., 2020, 2017b). As qPCR-based assays only detect a fragment of the genome, the question of virus integrity, and hence infectivity, remains open. Infectivity assays can offer a solution, but even where available, require specialised cultivation systems and are not likely to become generally applicable for

* Corresponding author. (E. Adriaenssens)
E-mail address: Evelien.adriaenssens@quadram.ac.uk (E.M. Adriaenssens).

https://doi.org/10.1016/j.watres.2021.117568
Received 24 May 2021; Received in revised form 10 August 2021; Accepted 11 August 2021
Available online 14 August 2021

routine monitoring of public health risks (DiCaprio, 2017). As a more comprehensive and now potentially feasible alternative, we applied shotgun viromics, i.e. next-generation sequencing of the entire aquatic virome, to reconstruct full virus genomes from the environment and objectively scrutinise the ecological and health implications of virus diversity and geographical distribution, with minimal bias.

Virome analyses are transforming our understanding of viral diversity and function in the biosphere (Emerson et al., 2018; Gregory et al., 2020; Roux et al., 2016) and provide unprecedented opportunity to understand the connectivity and fate of human-derived viruses at the catchment scale. Here, we present the first integrated analysis of the full virome of a river catchment system and estuary including water, sediment, wastewater treatment plants, beaches and shellfish production areas (Fig. 1 and Supplementary Table 1). We assembled over 40,000 partial or near-complete genomes (UViGs Uncultivated Virus Genomes, (Roux et al., 2019)) of ssRNA, dsRNA, ssDNA and dsDNA viruses, clustered into 10 149 species-level groupings (vOTUs, viral Operational Taxonomic Units). Our detailed bioinformatic analysis of the RNA and DNA viromes provides an assessment of viral diversity in the wastewater-impacted Conwy river catchment located North Wales (UK) (Farkas et al., 2018a, 2019; Perkins et al., 2014; Robins et al., 2019), encompassing information on the dynamics of viral deposition along the river system leading to viral enrichment at the estuary, including shellfish destined for human consumption and a recreational bathing beach. Viral genome reconstruction revealed general patterns of viral enrichment, dilution and degradation, and the implications for human health.

2. Materials and Methods

2.1. Sample collection

This work builds on our pilot study of a single wastewater treatment plant, and a downstream water and sediment sampling site, in which we optimised methods and showed that we could reconstruct RNA virus genomes from environmental samples (Adriaenssens et al., 2018). We collected and processed four different types of samples for this study: wastewater (influent and effluent), surface water (river and estuarine), sediment and shellfish in June 2017 from the Conwy river catchment area located in North Wales (UK) (Fig. 1, Supplementary Table 1). Wastewater influent was collected from the four major treatment plants in the catchment and the corresponding effluent from three (the Ganol plant effluent pipe exits directly into the open sea and therefore was not sampled separately); one litre per sample. Surface water was collected in four biological replicates of 50 L, resulting in two replicates per library type (RNA and DNA–based). At two locations, one in the river and one at a major recreational beach, sediment samples were taken in 4 biological replicates of at least 50 g, two replicates per library type. At low tide, samples were scooped from an accessible part of the river bank into sterile bags using a sterilised spade. Finally, mussels (*Mytilus edulis*) were collected from the two main commercial shellfishery locations in the estuary and divided into eight pseudoreplicates, as described below. Samples were collected at low tides, enabling the collection of shellfish samples from shore without a boat.

2.2. Sample processing

The wastewater (1 L) and surface water (50 L) samples were concentrated using a two-step protocol involving tangential flow ultrafiltration (TFUF) and beef extract elution as described in detail previously (Farkas et al., 2018c). Briefly, sample volumes were reduced to 50 mL by TFUF on a KrosFlo® Research IIi Tangential Flow Filtration System (Spectrum Labs, Phoenix, AZ, USA, Cat. no. SYR-U20-01N) using a 100 kDa cut-off mPES MiniKros® hollow fibre filter (Spectrum Labs). Virus particles were eluted using beef extract and $NaNO_3$ to a final concentration of 3% and 2 M (pH 5.5), respectively. After the solution's pH was adjusted to 7.5, PEG 6000 was added to a final concentration of 15% with 2% NaCl, incubated overnight at 4°C and after centrifugation (30 min, 10 000 x g, 4°C) the pellet was resuspended in 10-15 ml PBS (pH 7.4). These suspensions were kept at -80°C until nucleic acid extraction. The sediment samples were processed using beef extract elution and PEG 6000 precipitation as above and described previously (Farkas et al., 2017a).

For the mussel samples, approximately 200 mussels were collected from each location and stored on ice. Each mussel was dissected and the digestive tissue extracted and minced with a scalpel. The tissue was

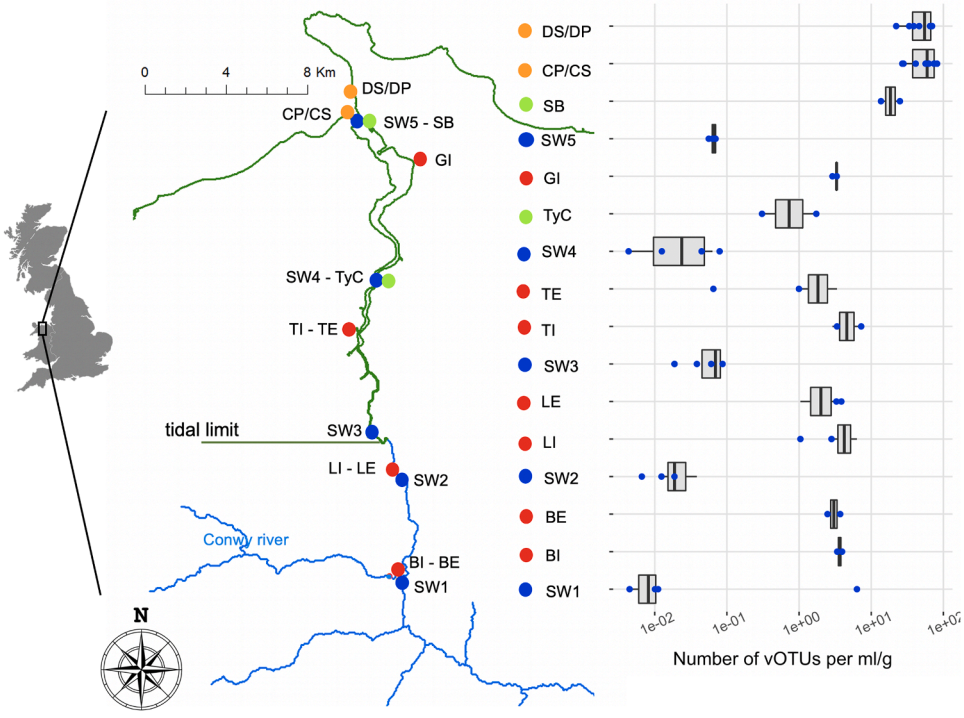

Fig. 1. Viral abundance along the wastewater impacted Conwy river catchment and coastal zone. Left: Schematic of the Conwy river catchment with sampling sites designated by colour-coded dots (red – wastewater, blue – surface water, green – sediment, orange – shellfish). The section of the river within the tidal limit is designated in green. Map of Great Britain by Free Vector Maps. Right: Boxplot representation of the number of species (vOTUs) detected in each sample per ml or g of sample extracted, composed of RNA and DNA libraries and biological replicates, species numbers for single libraries in blue dots.

pooled per location and then divided into four pseudoreplicates. Two replicates per location were mixed with SM buffer (0.1 M NaCl, 50 mM Tris/HCl–pH 7.4, 10 mM $MgSO_4$) and two with PBS at 25 g of digestive tissue to 20 mL of buffer. The samples were then shaken for 30 minutes (150 rpm) at room temperature to dissociate viral particles from the digestive tissue after which they were stored at -80°C. Due to a limit on the number of shellfish collected, we used both PBS and SM buffer as we could not perform a pre-optimisation of the protocol. No systematic differences between the two methods were detected.

2.3. RNA extraction

Wastewater, surface water and sediment concentrates were processed as follows. The concentrates were diluted in an equal volume 0.5 M NaCl to improve dissociation of viral particles before filtration. After centrifugation (5 min, 3200 x g) the supernatant was filtered through a 0.2 μm sterile syringe filter (Millipore). The filtrate was further concentrated using Vivaspin 20 spin filters (100 kDa) and centrifugation at 3200 x g. Once the volume was below 1 mL, 5 mL Tris buffer (5 mM TrisHCl, 5 mM $MgSO_4$, 75 mM NaCl, pH 7.5) was added and the volume reduced (two times) to reduce the NaCl content of the virus suspension. Centrifugation times ranged between 150 minutes and 20 hours to reduce the volume below 500 μL for the next step. A DNase treatment with 10 U Turbo DNase (Invitrogen) was performed to remove extra-viral DNA (incubation at 37°C for 30 min, inactivation at 75°C for 10 min). Mussel samples were highly viscous and required separate processing, as we were unable to filter or concentrate with the Vivaspin filters. Instead, 2 × 1 mL aliquots per replicate were mixed with 0.1 mm glass beads (MoBio) and lysed in a PowerLyser (MoBio) shaker (2 × 30 seconds at 3400 rpm). Debris was removed by centrifugation (5 min at 3200 x g) and the supernatant was stored at -20°C for next-day processing.

For all sample types, the viral capsids were lysed using a combination of proteinase K (50 μg for clear samples, 100 μg for turbid samples), EDTA (0.5 M final concentration) and SDS (0.5% final concentration), and incubation for one hour at 56°C. Next, the RNA was extracted by TRIzol extraction derived from Kroger and colleagues (Kroger et al., 2012). In short, 500 μL of sample was mixed with 1 mL of TRIzol reagent and 200 μl of molecular-grade chloroform in Phasemaker™ tubes (Invitrogen), shaken vigorously and centrifuged for 15 minutes at 13 000 x g. The aqueous phase was removed and transferred to a new tube. The phase separation was repeated for samples that remained turbid. The nucleic acid was recovered by isopropanol precipitation and resuspended in 50 μL of sterile, RNase-free water. Viral DNA was removed with an additional DNase step, adding 4 U Turbo DNase, 5 μL TD buffer, and incubating for 40 minutes at 37°C followed by inactivation of the DNase at 75°C for 10 minutes. The DNase was removed by a second isopropanol precipitation as above, the RNA resuspended in 50 μL of RNase-free water and stored at -80°C until sequencing. Alongside all samples, a positive extraction control comprising of *Salmonella* cells (*Salmonella enterica* subsp *enterica* serovar Typhimurium strain D23580, RefSeq acc NC_016854) and the process-control virus mengovirus (~ 10^5 particles/ml) was extracted, as was a negative Tris buffer control.

2.4. DNA extraction

Wastewater, surface water and sediment samples were processed similarly as for RNA extraction with a few amendments to the extraction process. The samples were diluted in 10 ml NaCl (0.5 M). For surface water and sediment samples one replicate (designated a) was treated with chloroform (1 mL) to lyse the cellular fraction (15 min incubation with gentle shaking) and the cellular debris removed by centrifugation (5 min, 3200 x g). The second replicate (b) was filtered through a 0.45 μm sterile syringe filter (Millipore). For the wastewater samples which consisted of only one replicate, the sample was split in two, half treated with chloroform and half filtered, and then merged. All samples were then concentrated and desalted as described above (using Vivaspin 20 spin filters (100 kDa) and centrifugation at 3200 x g, with centrifugation time between 100 min and 20h). All sample concentrates (approx. 500 μL each) were treated with 10 U of Turbo DNase (Invitrogen) and 10 μg of RNase A (Thermo Fisher Scientific) supplemented with Turbo DNase buffer for 30 min at 37°C and inactivation at 65°C for 10 min. The separation of filtering and chloroform treatment before proteinase and nuclease treatment theoretically allows for the recovery of both giant viruses (larger than the filter pore size) and viruses with lipid membranes, while simultaneously reducing the background cellular DNA. No differences between treatments were found for potentially pathogenic viruses, however, the systematic comparison of methods for all viruses is beyond the scope of this study and we encourage others to use our data for further exploration.

Mussel digestive tissue was processed exactly as during RNA extraction (mixed with 0.1 mm glass beads and lysed in PowerLyzer) and no nuclease treatment was performed.

From this point, all samples were extracted in the same manner. Capsids were lysed by adding proteinase K (50 μg/mL final concentration), EDTA (20 mM final concentration) and SDS (0.5% final concentration), followed by incubation at 56°C for one hour and extracted using phenol/chloroform/isoamylalcohol (25:24:1) in PhaseLock tubes (VWR). The aqueous phase was transferred to a new tube and the process was repeated at least once (twice for turbid samples), followed by one round of chloroform phase separation. Finally, samples were further cleaned and concentrated with ethanol precipitation (2.5 x volume 100% ethanol; 1/10 volume 3 M NaAc pH 5; incubation at -20°C for 30 minutes; precipitation 30 min at 15 000 x g, 4°C), washed with 70% ethanol and air-dried in a laminar flow cabinet.

In tandem with the whole process, control samples were extracted, starting with the dilution in NaCl. We used a negative control consisting of Tris buffer and a positive control consisting of 500 μl stationary culture *Escherichia coli* MG1655 cells (RefSeq acc NC_000913), 2.2 × 10^8 pfu of Escherichia phage T5 (RefSeq acc NC_005859) and 1.3 × 10^5 pfu of Escherichia phage vB_EcoP_phi24B (GenBank acc HM208303).

2.5. Sequencing

Sequence library preparation and sequencing was performed by the Centre for Genomics Research (CGR) NBAF facilities at the University of Liverpool, UK. RNA libraries were prepared as in the pilot study (Adriaenssens et al., 2018) using the NEBNext Ultra directional RNA library preparation kit of Illumina with dual indexes. During library preparation, the number of PCR cycles was increased to 30 to account for the low amounts of input RNA (< 1 ng). Dual-indexed DNA libraries were generated using the NEBNext Ultra II DNA Library Prep kit according to the manufacturer's instructions. Libraries were pooled and sequenced on six lanes of the Illumina HiSeq 4000 generating paired-end 2 × 150 bp reads, three lanes for the RNA libraries in July 2017 and three lanes for the DNA libraries in March 2018.

The DNA sequencing run failed for libraries DNA_SW2a, DNA_TyCa/b, DNA_SBa/b and unfortunately, we were unable to reconstruct the libraries as the samples had been mistakenly stored at 4°C and the DNA had degraded. Furthermore, the read lengths obtained for the mussel DNA libraries were much lower as for all other libraries, as the DNA had been excessively sheared during the extraction procedure.

2.6. In silico processing

Reads went through an initial round of quality control at CGR to remove Illumina adapters (Cutadapt version 1.2.1, -O 3) and were trimmed with Sickle (version 1.2) removing all reads below an average quality of 20 and shorter than 20 bp (Joshi and Fass, 2011; Martin, 2011). The resulting fastq files were received as raw read files from the CGR and deposited into SRA under BioProject PRJNA509142, accession numbers SRR8299359 to SRR8299398.

The paired-end read files were further trimmed and filtered to increase quality using the prinseq-lite suite (Schmieder and Edwards, 2011) and the read pairs meeting the following criteria were retained: minimum length 35 bases, GC-content between 5 and 95%, maximum 1 N, trimmed until the average read quality was 30. For all exactly duplicated reads only one copy was retained. The reads for the control libraries were merged per library type (RNA & DNA) and used as a bowtie2 mapping reference (Langmead and Salzberg, 2012). Each of the sample libraries was then mapped against its control and only the unmapped reads were retained. These reads were then assembled per sample using SPAdes version 13.9 using the k-mers lengths 21,33,55,77, 95,107,121 (Nurk et al., 2013), with the exception of the mussel DNA libraries containing the shorter reads where the k-mers 21,33,55,77 were used. The control libraries were assembled using the same parameters and compared to the sample contigs using BLASTn (BLAST+ suite), and sample contigs that showed significant similarity (e value $<$ 0.001) were removed from each of the sample contig datasets (Camacho et al., 2009).

From these contigs, an Anvi'o contig database was created according to the instructions of the metagenomics workflow (Eren et al., 2015). To be included in the database, contigs needed to meet the following criteria: RNA library assemblies (i) contig length min 1000 nucleotides (nt); (ii) amino acid similarity with any known virus; (iii) recruit no reads from control libraries; DNA library assemblies (i) contig length min 10,000 nt, (ii) identified by VirSorter as viral in categories 1 or 2 (Roux et al., 2015), (iii) recruit no reads from control libraries. VirSorter was run on all DNA contig sets using the microbiome decontamination mode on the iVirus Cyverse infrastructure (Bolduc et al., 2017). The contig dataset comprising 40 000 UViGs was merged and clustered at an approximation of the viral species level (95% average nucleotide identity over min 80% of contig length), according to the species definition for bacteriophages implemented by the International Committee on Taxonomy of Viruses (ICTV) and conventionally used in virome studies (Adriaenssens and Brister, 2017; Emerson et al., 2018; Gregory et al., 2016; Roux et al., 2019, 2016). We performed a final refinement by removing all contigs $<$ 10 000 nt assembled from RNA libraries that showed amino acid similarity with dsDNA viruses, based on diamond BLASTx comparison (Buchfink et al., 2015) with the nr database downloaded from the NCBI in January 2018. The final database contained 10 149 UViGs (Uncultivated Viral Genomes, (Roux et al., 2019)) that each represent a viral species-level population. Taxonomic information was added to the contigs database in Anvi'o using Kaiju with the built-in viral database (Menzel et al., 2016). All individual assemblies were also compared with the nr and viral RefSeq protein databases (version Jan 2018) separately in case the length thresholds for contig database creation excluded certain virus types.

To compare the incidence and abundance of UViGs in the different samples, for each library the reads were mapped to the contigs database using kallisto (Bray et al., 2016). The abundances of contigs within and between samples were assessed by transforming the values into Transcript Per Million values (TPMs) where each contig (UViG) was considered a transcript using the program tximport in R (Soneson et al., 2016). The resulting 10 149 by 58 matrix was visualised with Phantasus (Zenkova et al., 2018). The pseudobam alignment files generated by kallisto were then transformed into Anvi'o profiles according to the metagenomics workflow instructions and investigated using the anvi-interactive interface (Eren et al., 2015). Numbers of species detected per library, sample or sample type were calculated as the number of UViGs having a TPM value of minimum 10. Venn diagrams were produced on the online webserver http://bioinformatics.psb.ugent.be/webtools/Venn/ hosted by the VIB-UGent Center for Plant Systems Biology.

The taxonomic classifications by Kaiju as part of the Anvi'o platform left over 5000 UViGs unclassified. We then used diamond BLASTx against the viral RefSeq protein database (version 200, May 2020) and Megan 6 Community Edition to assign all UViGs to their most reliable taxonomic rank using the Megan 6 "long read" lowest common ancestor algorithm at the default settings (Buchfink et al., 2015; Huson and Weber, 2013). The taxonomic bin information was added to the Phantasus heatmaps by matching the UViG names and exported to R Studio using the tidyverse packages to create graphs in ggplot2 (Wickham, 2016; Wickham et al., 2019).

To generate phylogenetic trees of taxonomic groups of interest, we used the Megan 6 taxonomic bins. All UViGs assigned to a bin were annotated with Prokka (Seemann, 2014) using the -kingdom Viruses setting and the predicted CDSs were manually curated in UGene (Okonechnikov et al., 2012) to adjust for the presence of polyproteins and missing start or stop codons from incomplete genomes. Per RNA virus taxonomic group, the RNA-dependent RNA polymerase (RdRP) amino acid sequences were extracted and aligned together with RdRP sequences from reference databases using MAFFT with maximum 5 iterations (Katoh and Standley, 2013). The resulting alignments were trimmed with TrimAl (Capella-Gutiérrez et al., 2009) using the -gappyout setting, followed by manual inspection in the UGene alignment viewer. Sequences missing the conserved structural motifs present in RdRPs (Venkataraman et al., 2018) were removed, as were sequences missing more than 50% of the trimmed sites. Trees were computed using the IQ-Tree suite (Nguyen et al., 2015) including calculation of the best substitution model with ModelFinder (Kalyaanamoorthy et al., 2017), calculation of the approximate likelihood ratio test (1000 repetitions) (Anisimova et al., 2011) and ultrafast bootstrap approximation with UFBOOT2 (1000 repetitions) (Hoang et al., 2018). The resulting trees were analysed and annotated in iTOL (Letunic and Bork, 2007). For the picorna-calici tree, the alignments generated by Shi and colleagues were additionally used as references (Shi et al., 2018, 2016).

Rotavirus segment genotyping was performed on the RotaC 2.0 webserver of the Rega Institute (KU Leuven, Belgium) (Maes et al., 2009).

3. Results and Discussion

3.1. Viral species richness is environment-specific and geographically distinct

We generated a final species-level contig database containing 10 149 vOTUs from 44 897 assembled contigs, each represented by the longest viral genome (UViG). We used the number of vOTUs per sample, normalised per volume (ml) or wet weight (g) of input material, as an approximation of species richness (Fig. 1). Normalised viral OTU ("species") richness was highest in the shellfish digestive tissue and beach sediment samples, intermediate in wastewater influent and effluent and lowest in the surface water samples. The surface river water samples showed a trend of increasing viral richness moving downstream, as further inputs of wastewater from treatment plants and other anthropogenic sources occurred, reaching a plateau around the location of SW3 after which richness remained high (Fig. 1).

We further investigated the differential patterns of abundance of each UViG in each library by mapping reads of all samples against the vOTU database and visualised the data with Anvi'o (Fig. 2, Supplementary Fig. 1), to identify 13 categories of viral species abundance and composition patterns (Table 1). The wastewater samples contained the most diverse set of UViGs in absolute numbers, however, each wastewater treatment plant yielded a geographically distinct viral community. The river water samples contained a lower absolute richness of viruses than the wastewater, except for sample SW5 (and to a lesser extent, SW3) which showed a high degree of overlap in UViGs from category 1 with the wastewater influent sample from the Tal-y-Bont treatment plant (RNA_TI) (Fig. 2). Many of the same UViGs in this category (1) were also detected in the mussel (shellfish) samples and in the sediment samples. Comparing this pattern with the geographical origin of the samples (Fig. 1), revealed that the river water upstream and distant from wastewater effluent locations contained fewer detectable

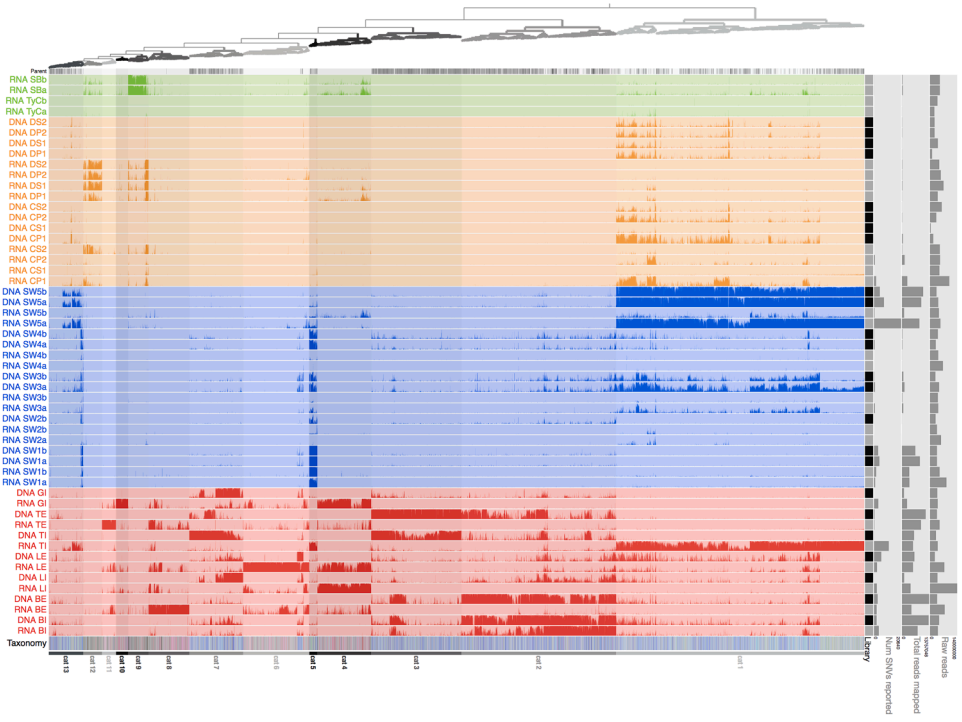

Fig. 2. Differential patterns of abundance of each viral genome (UViG) along the wastewater impacted Conwy river and coastal zone. Anvi'o - mean coverage per contig (split). Each row is a sequencing library, coloured by its sample type (green = sediment; orange = mussels; blue = river/estuary water; red = wastewater). Each column (leaf in top dendrogram) is a contig or a split of a contig (in cases where contigs were larger than 11 kb). The height of the bar in each row is the log mean coverage across the contig or contig split length. The contigs are clustered (top dendrogram) according to their sequence composition and differential coverage using Euclidean distance and Ward linkage. Based on this clustering, we identified 13 categories of UViGs, indicated by shades of grey in the dendrogram and numbered at the bottom of the plot. The bottom row represents the taxonomy assigned by Kaiju (using its viral database) to the predicted genes in each contig. Contigs without assigned taxonomy are depicted in grey, dsDNA bacteriophages in shades of blue, other dsDNA viruses in shades of green, ssDNA viruses in shades of yellow, RNA (ds, (+)ss, (-) ss) in shades of purple/red. The right hand panels show the library type (RNA = grey; DNA = black), the number of single nucleotide variants (SNVs) found after read mapping (0 - 20 640), the total number of reads mapped to contigs (0 – 13 757 048) and the total number of raw sequencing reads (before QC and contamination screen; 0 – 140 000 000).

Table 1
Categories of UViGs observed in the dataset, binned using a combination of sequence composition and read mapping pattern.

Groups	Number of UViGs	Total length (Mb)	N50 (nt)	Sample presence	Main virus types per category
cat_1	3257	35	18100	WW, SW, SF	dsDNA phages, NCLDVs[a]
cat_2	1514	27	29090	WW, SW	dsDNA phages, NCLDVs
cat_3	636	18	38076	WW, SW	dsDNA phages, NCLDVs
cat_4	890	1.9	2446	WW, SW, SF, Sed	(+)ssRNA phages, dsRNA viruses, RNA plant viruses
cat_5	103	1.3	17154	WW, SW	dsDNA phages, NCLDVs
cat_6	1077	3.9	5239	WW, SW, SF	(+)ssRNA viruses, dsRNA viruses, dsDNA phages
cat_7	519	11	24406	WW	dsDNA phages, NCLDVs
cat_8	671	2.1	3765	WW	(+)ssRNA viruses, ssDNA viruses
cat_9	337	1.0	4133	SF, Sed	(+)ssRNA viruses, dsRNA viruses, ssDNA viruses
cat_10	200	0.36	1750	WW	(+)ssRNA viruses, dsRNA viruses, ssDNA viruses
cat_11	230	0.65	3703	WW	(+)ssRNA viruses, ssDNA viruses
cat_12	309	0.88	3959	SF, Sed	(+)ssRNA viruses, dsDNA phages
cat_13	406	6.2	20764	WW, SW, SF	dsDNA phages, NCLDVs

[a] NCLDV: nucleo-cytoplasmic large DNA virus. Key WW – wastewater, SW – surface water, SF – shellfish, Sed – sediment.

virus species, while the locations immediately downstream of an effluent pipe (SW3) or in the tidal estuary (SW5, mussels, beach sediment) were enriched for UViGs. The high virus richness in the tidal estuary (SW5) can be explained by the mixing of river and marine waters during tidal flow (Robins et al., 2019). The SW5 wastewater treatment Combined Sewage Outflow (CSO) periodically discharges untreated sewage directly upstream of SW5, representing a sewage input that largely avoids the dilution effect of estuary water and is consistent with the higher detection of faecal indicator bacteria previously at this location (Perkins et al., 2014). The viral species count per sample (Fig. 1) also demonstrated that wastewater effluent samples (mean 1287) generally had a lower species tally than influent (mean 2079), indicating that wastewater treatment reduced viral species richness, but the large variance and limited number of samples (n = 4 per group) did not allow for meaningful tests of statistical significance.

Examination of UViGs grouped per environment type for shared viral species [cut-off for detection 10 TPM (transcripts per million, see methods)] confirmed our observation that absolute richness was highest in wastewater samples (2692 unique vOTUs; Fig. 3a). River/estuarine water (82 unique UViGs), mussels (137) and sediment (100) all contained an order of magnitude fewer unique vOTUs. The majority of the vOTUs present in mussels and sediment were shared with wastewater; out of 4692 vOTUs detected in mussels, 3917 were also detected in wastewater (83%), and for sediment this was 1464 out of 1944 (75%) (Fig. 3a). Even though most of these vOTUs were likely bacteriophages, the high connectivity of these environments is a cause for concern, indicating potential sources of contamination that pose a risk to human health, and is investigated in more detail below. The categories of vOTUs in wastewater encompassed all virus types detected in this study (Table 1), whereas those specific to mussel shellfish and sediment comprised primarily RNA viruses. In the Materials and Methods Sequencing section, we describe technical difficulties during sequencing library construction that may have resulted in the underestimation, or even failure to detect, a group of mussel and sediment-specific DNA

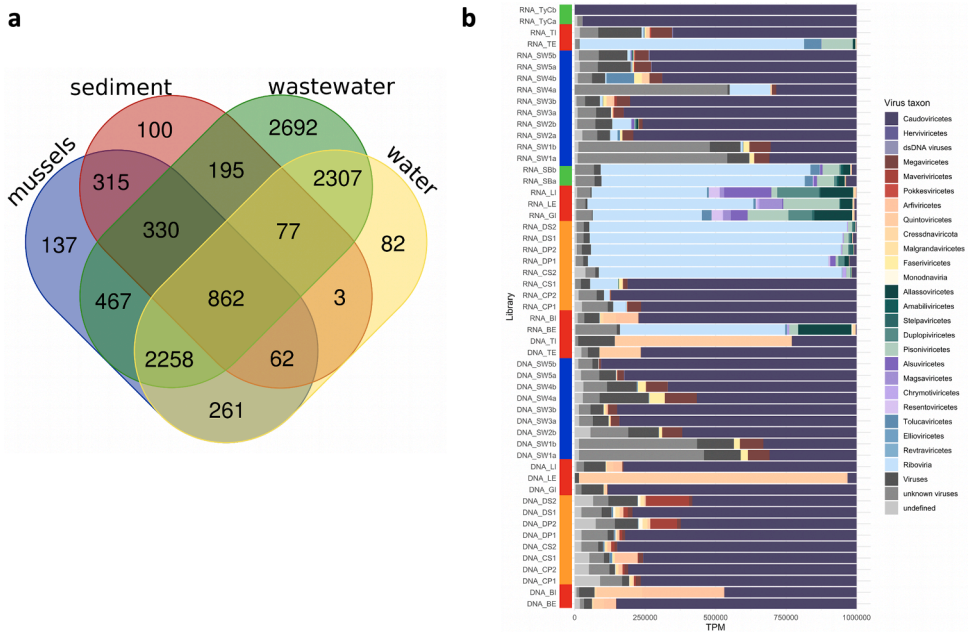

Fig. 3. Commonality and taxonomic composition of viral genomes (UViG) in samples types from the wastewater impacted Conwy river and coastal zone. a, Venn diagram representation of the number of UViGs shared between different environment types (min 10 TPM for detection). b, Relative abundances of the UViGs at the virus class level per sequencing library (colour-coded per sample type as in figure 2, green = sediment; orange = mussels; blue = river/estuary water; red = wastewater) normalized per library as transcipts (=contig) per million (TPM). dsDNA viruses in shades of dark purple and red; ssDNA viruses in shades of pink and yellow; RNA viruses in shades of green, purple and blue; unknown viruses in shades of grey.

viruses.

In order to assign each UViG to a viral family and higher taxa, we used a combination of Diamond BLASTx against the viral RefSeq protein database (version 200, May 2020) and taxonomic binning using a lowest common ancestor approach with Megan6 (Buchfink et al., 2015; Huson and Weber, 2013). To reduce the number of different taxa displayed in Figure 3b, we assigned the UViGs at class or phylum level recently defined by the International Committee on Taxonomy of Viruses (ICTV) (Gorbalenya et al., 2020; Koonin et al., 2020) and where this was not unambiguously possible at the Realm level, with the remainder designated as either "Viruses" (similarities to viruses belonging to multiple Realms) or "Unknown" (no similarity with any virus in the RefSeq database). Of all contigs in our set, 98% had at least one BLAST hit with the virus database (9935/10 149) and 88% (8904/10 149) were assigned to at least the viral Realm level.

The taxonomic composition of each library (Fig. 3b), normalised to 1 million reads per sample mapped to the UViGs, showed large differences in relative abundances of virus groups between both library types and samples. Scanning these data confirms the observation from Figure 2, that some of the RNA libraries were contaminated with DNA viruses. In these cases (all RNA river water libraries and RNA_TI, RNA_BI, RNA_TyCa/b, RNA_CP1/2/CS1), the relative abundance of dsDNA viruses, mainly tailed phages of the class *Caudoviricetes*, eclipsed the detected RNA virus signatures. The remainder of the RNA libraries recruited the most reads against several groups of RNA viruses, such as the phage class *Leviviricetes* (formerly family *Leviviridae*), phylum *Lenarviricota* and unknown RNA virus UViGs (Realm *Riboviria*) from a previously published study on the RNA virosphere of invertebrates (Shi et al., 2016). The majority of the DNA libraries were dominated by dsDNA bacteriophages associated with the class *Caudoviricetes* and its constituent families. Exceptions were libraries DNA_TI and DNA_LE, which were dominated by a small number of UViGs with ambiguous taxon assignments (i.e. classified as "Viruses" or "Unknown"). The read recruitment to the UViGs and their taxonomic binning clearly showed discrepancies between some of the replicates, most notably the RNA libraries of the shellfish digestive tissue samples. These differences are in line with a recent study by Pérez-Cataluña and colleagues who investigated library preparations for viromes of wastewater and showed that further standardisation of methods is necessary for quantitative viromics (Pérez-Cataluña et al., 2021).

In view of these discrepancies in read mapping patterns between replicates, we investigated the taxonomic bins per environment type as an indication for richness, not relative abundance (Supplementary Fig. 2). Overall, the most common RNA virus classification was the "UViG RNA virus" bin grouped within the realm *Riboviria* comprising a diverse set of metagenome-assembled RNA viruses [dsRNA, (+)ssRNA, (-)ssRNA from invertebrates (Shi et al., 2016)], which contained the most UViGs from mussels, sediment and wastewater samples. The most abundant DNA virus bin was the class *Caudoviricetes* which groups all tailed phages of the order *Caudovirales* and its constituent families (*Myoviridae, Siphoviridae, Podoviridae, Ackermannviridae, Autographiviridae, Drexlerviridae, Herelleviridae*) including crAss-like phages, and unidentified dsDNA viruses (probably tailed bacteriophages), which were particularly rich in wastewater, river water, and to a lesser extent in mussels. Wastewater was also host to a diverse group of (+)ssRNA phages of the class *Leviviricetes* (~700 UViGs), with a smaller number of these viruses observed in mussels and sediment. About 6% of the total reads could not be assigned to a known group, not even at the Realm level, and were categorized as unknown viruses. While these unknown viruses represented only 6% of the total reads, they made up about a third (3 502/10 149) of the vOTUs.

3.2. *Circulating human pathogens: Sapovirus, coxsackievirus and rotavirus*

To investigate the potential environmental and public health impact of the UViGs, we focused on the near-complete genomes shared between wastewater and the other environments (Fig. 3a) and the taxonomic groups that contain known pathogens (human/animal). We identified 29 vOTUs of potential public health concern, further representing 73 UViGs from six families (Table 2). Interestingly, we were unable to unambiguously identify any potentially pathogenic dsDNA UViGs. The ability to reconstruct a complete papillomavirus genome in our pilot study from a subset of these sites sampled at an earlier date (Adriaenssens et al., 2018) suggests that there were in fact no predicted-pathogenic DNA viruses circulating (above the limit of detection) in the Conwy catchment at the time of sampling (June 2017). We speculate that the most likely reason for the absence of potentially pathogenic DNA viruses from the dataset is because their presence was below the limit of detection, with the discovery of the papillomavirus in the pilot study potentially due to a large shedding event. It is also possible that the diversity of dsDNA bacteriophages drives up the limit

Table 2
Potentially pathogenic virus groups in the UViG dataset.

Family/group	Genus – closest relative	Potential host/ metagenome	# of contigs[a]	Cat[b]	Present in samples (traces)[c]
Astroviridae	UviG Bastro-like virus*	Bat	1	12	DM
(+)ssRNA	*Astrovirus* - Astrovirus MLB1	Human	1	10	GI
Caliciviridae (+)ssRNA	*Sapovirus* - Sapovirus GII.5	Human	6	4	LI, LE, GI (SB, DM, SW5)
	Sapovirus - Sapovirus GII.2	Human	2	4	LI, LE (GI, SB, DM, SW5)
Picornaviridae (+)ssRNA	*Enterovirus* - Human coxsackievirus A22	Human	1**	10	GI (SB)
	Enterovirus - Human coxsackievirus A19	Human	1**	10	GI (SB)
Reoviridae	*Rotavirus* - Rotavirus A (NSP1)	Human	2	4	LE, GI (LI, SB, DM)
dsRNA	*Rotavirus* - Rotavirus A (VP1)	Human	2	4	LE, GI (LI, SB, DM)
	Rotavirus - Rotavirus A (VP2)	Human	2	4	LE, GI (BI, LI, SB, DM)
	Rotavirus - Rotavirus A (VP3)	Human	2	4	LI, LE, GI (SB, DM)
	Rotavirus - Rotavirus A (NSP1)	Human	4	4	BE, LI, LE, GI
	Rotavirus - Rotavirus A (NSP3)	Human	4	4	BE, LI, LE, GI (TE, SB, DM, SW5)
	Rotavirus - Rotavirus A (VP1)	Human	3	4	BE, LI, LE, GI (TE, SB, DM, SW5)
	Rotavirus - Rotavirus A (VP2)	Human	4	4	BE, LI, LE, GI (TE, SB, DM, SW5)
	Rotavirus - Rotavirus A (VP3)	Human	4	4	BE, LI, LE, TE, GI (SB, DM, SW5)
	Rotavirus - Rotavirus A (VP4)	Human	4	4	BE, LI, LE, GI (TE, SB, DM, SW5)
	Rotavirus - Rotavirus A (VP7)	Human	4	4	BE, LI, LE, GI (TE, SB, DM, SW5)
	Rotavirus - Rotavirus A (NSP1)	Human	1	6	LE
	Rotavirus - Rotavirus A (NSP3)	Human	1	6	LE (LI)
	Rotavirus - Rotavirus A (VP1)	Human	1	6	LE (LI)
	Rotavirus - Rotavirus A (NSP3)	Human	1	10	GI (LI, LE, SB)
	Rotavirus - Rotavirus A (VP1)	Human	1	10	GI (LI, SB, DM)
	Rotavirus - Rotavirus A (VP4)	Human	1	10	GI (LE, SB)
	Rotavirus - Rotavirus A (VP7)	Human	1	10	GI
Circoviridae	UviG CRESS-like virus	Animals	8	1	BE (LI, LE, TI, SW4)
ssDNA	UviG CRESS-like virus	Animals	1	2	SW5, TI
	UviG Human fecal virus Jorvi3	Human	2	2	SW3, TI, LI, BI, BE
	UviG Giant panda circovirus 1	Mammals	7	2	TI, TE
Parvoviridae ssDNA	*Ambidensovirus* - Densovirus SC444	Bat	1	6	LE (SW3)

*This assignment was based on low similarity scores.
**These UViGs were partial genomes, not near-complete genomes.
[a] The number of UViGs clustered at 95% ANI (cd-hit-est) represented by one UViG in the dataset.
[b] Category as defined in Fig. 2 and Table 1.

of detection for other DNA viruses by skewing the data towards the most abundant genomes. Given the current pandemic, we also did a search for similarity of the UViG dataset with members of the *Coronaviridae* family, but no coronavirus signatures were identified in our dataset.

With respect to the family *Astroviridae*, we recovered one UViG related to bat-infecting astroviruses in mussel (*Mytilus edulis*) tissue from the Deganwy shellfishery, and one UViG in wastewater sample GI highly similar to Astrovirus MLB1 (FJ222451) which was sequenced from the stool of a child with acute diarrhoea (Finkbeiner et al., 2008) (Supplementary Fig. 3a). For the family *Caliciviridae*, we were not able to identify any UViGs representing noroviruses, the leading cause of viral gastro-intestinal illness in the UK and indeed worldwide (Ahmed et al., 2014; FSA, 2017; Kirk et al., 2015) in contrast to our pilot study performed in autumn, where we assembled a norovirus GI.2 genome (Adriaenssens et al., 2018). We did, however, find two near-complete sapovirus UViGs and six shorter contigs grouped with the near-complete genomes (Fig. 3, Supplementary Fig. 3b), most closely related to sapoviruses of genotype GII.2 and GII.5 that were collected from children with acute gastroenteritis in Nashville (US) (Diez-Valcarce et al., 2018). This finding suggests that at the time of sampling for the dataset reported here (June 2017), sapoviruses replaced noroviruses (commonly associated with winter illness) as the main cause of gastro-intestinal disease. This theory is supported by our previous RT-qPCR detection study showing that sapovirus concentrations spiked between March and June in wastewater collected at the four WWTPs in the Conwy area (Farkas et al., 2018a). However, this is difficult to formally prove as many norovirus-sapovirus cases are undiagnosed clinically, and the seasonality of norovirus and sapovirus is not consistent in all clinical settings in the UK (Brown et al., 2016; Inns et al., 2019).

We identified two potentially pathogenic UViGs in the *Picornaviridae* family (Table 2) among a host of distantly related picorna-like viruses (Fig. 4). The two potentially pathogenic picornavirus UViGs, which were represented by only partial genome sequences (Supplementary Fig. 3c), could be identified as coxsackieviruses of the species *Enterovirus C*, most closely related to human coxsackieviruses A19 and A22 reportedly involved in meningitis, gastroenteritis and herpangina (Tapparel et al., 2013; Zell et al., 2017). Detailed phylogenetic analysis of all calici- and picorna-like RNA-dependent RNA polymerase (RdRP) sequences (Fig. 4) showed that the majority of UViGs found in this study fell within a very diverse, ill-resolved clade (low branch support) comprised of environmental sequences nested within the order *Picornavirales* (bottom half of circle, Fig. 4). Based on the RdRP sequences, only three UViGs in the picorna-calici group were designated potential human pathogens (Fig. 4, black arrows), the two sapovirus UViGs and one of the coxsackievirus UViGs. Only the sapovirus UViG LI_NODE_9 was detected in all sample types, posing a potential risk for human health as it was detected in the mussel beds of the commercial shellfishery, sediment on the tourist beach and estuarine water (Supplementary Fig. 4). PCR-based detection of sapoviruses in older studies show that among cases of gastro-intestinal disease, sapoviruses accounted for only 4% of cases (vs 36% for noroviruses) (Amar et al., 2007), however, the primers used in that study (SR80 (Noel et al., 1997), JV33 (Vinjé et al., 2000)) did not match the two sapovirus genomes reconstructed in this study (data not shown). The detection of this complete genome sequence from two different wastewater treatment plants is another indication that sapoviruses are more common in the UK than previously reported, similar to its incidence reported in other countries (Mann and Liebert, 2019; Pang et al., 2019; Varela et al., 2018).

While the phylogenetic analysis does not provide enough evidence for the presence of plant-pathogenic picorna-like viruses in the Conwy river catchment, there is a set of UViGs present that is mollusc-specific (coloured orange in Figure 4). It is therefore likely that we have sequenced and reconstructed a set of mussel/shellfish-associated or –infecting viruses.

Within the non-redundant, species-level clustered dataset, 18 UViGs grouped into three categories according to read recruitment pattern, and were assigned to the species *Rotavirus A* in the family *Reoviridae*, representing a further 41 contigs. Analysis of reoviruses is confounded by their segmented nature, i.e. members of the genus *Rotavirus* contain 11 segments of dsRNA, and the size of the smaller segments is below our 1000 nt contig length threshold. We therefore analysed all contigs larger than 500 nt for the presence of rotavirus signatures and assigned genotypes to each segment recovered (Fig. 5a). The most common rotavirus A (RVA) genome constellation recovered was G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2, with additional genotypes R1 for the RNA-dependent RNA polymerase (RdRP) segment, C1 for the segment encoding VP2, P[1] and P[14] for the outer capsid-encoding segment, A3 and A11 for NSP1 and T6 for NSP2. In many of the wastewater samples, we assembled multiple contigs of the same segment indicating the presence of several co-circulating population lineages of rotavirus A in the population. Phylogenetic analysis of the outer capsid proteins (VP4) confirmed the genotype clustering, and comparison with isolated rotavirus VP4 sequences points towards a human origin for the P[4] and P[14] genotypes (found in samples BE, LI, LE and GI) and a potential bovine zoonotic origin for the P[1] genotype segment (Fig. 5b). The RVA genome segments recovered here are markedly different to those recovered from wastewater influent from Llanrwst (LE-LI) in our pilot study 10 months prior (Adriaenssens et al., 2018), for which the dominant genotypes of RVA were G8/G10-P[1]/P[14]/P[41], and a diverse set of rotavirus C segments were also present. We can conclude that rotavirus shedding into wastewater within the population varied both spatially and temporally, but more data are required to investigate any possible seasonal patterns. From the distribution of the rotavirus fragments in shellfish, beach sediment and estuarine water (Table 2), we speculate that rotaviruses could pose a potential risk for human health in relation to shellfish consumption or recreational activities and bathing within the immediate coastal zone. However, rotaviruses mainly affect

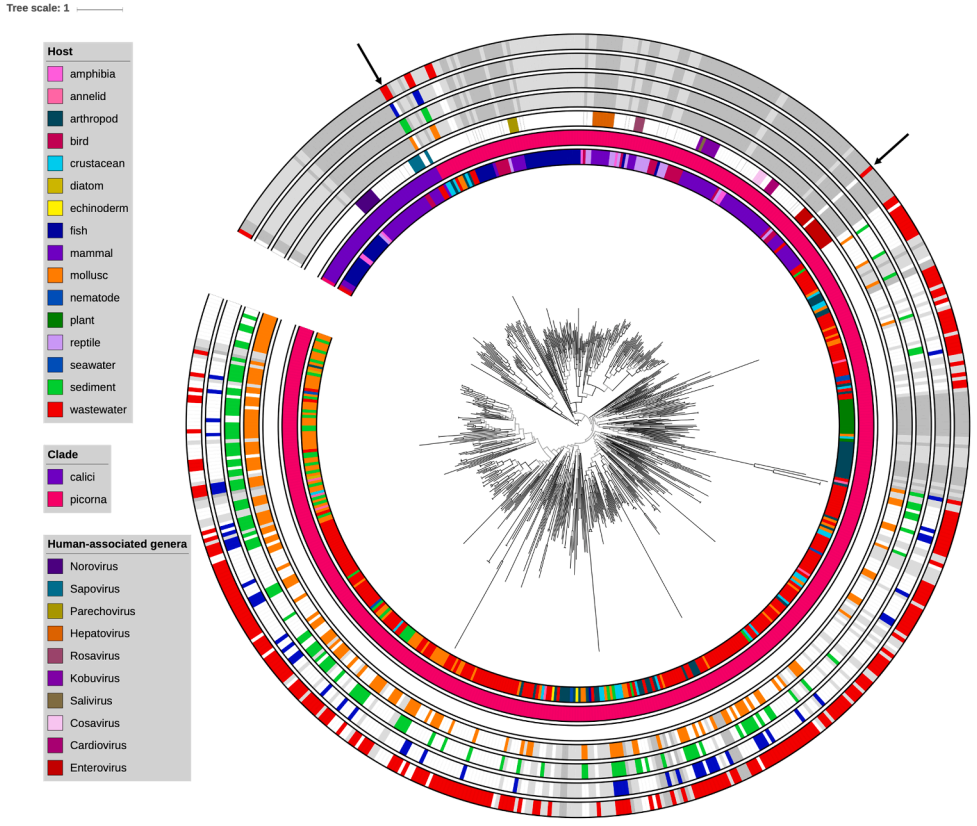

Fig. 4. Maximum likelihood phylogenetic tree of the RdRP amino acid sequences of viruses/genomes assigned to the family *Caliciviridae* and the order *Picornavirales* built with IQ-TREE (Nguyen et al., 2015) and visualized with ITOL (Letunic and Bork, 2019). The multiple alignment consisted of 622 sequences and 695 amino acid sites, aligned using MAFFT and trimmed with Trimal (Capella-Gutiérrez et al., 2009; Katoh and Standley, 2013). The best fit model was LG+F+R10 as determined with ModelFinder (Kalyaanamoorthy et al., 2017). Branch support was calculated using the Shimodaira Hasegawa – approximate Likelihood Ratio Test (SH-aLRT) and the UFBoot (ultrafast bootstrap) algorithm on 1000 replications with nodes below 80% (SH-aLRT) and 95% (UFBoot) indicated in grey (Anisomova and Gascuel, 2006; Hoang et al., 2018). The three inner colour strips from inside to outside indicate respectively: viral host or metagenome the RdRP was extracted from, predicted clade, human-associated genera (only reference genomes from human pathogenic viruses coloured). The four outside colours strips indicate detection in shellfish samples (orange), beach/river sediment samples (green), river/estuarine water samples (blue) and wastewater samples (red), with other virome-derived UViGs in light grey and reference virus sequences in middle grey. The black arrows indicate the UViGs found in this study that are likely human pathogens.

Fig. 5. Rotavirus A (RVA) in the virome datasets. a, The 11 segments of the reference genome of RVA ranked by size in black. RVA segments recovered per sample below showing the predicted genotype of the segment and the percentage of nucleotide identity with a representative of that genotype as calculated by the RotaC 2.0 tool. b, Maximum likelihood phylogenetic tree of the VP4 amino acid sequences of selected representatives of all RVA genotypes build with IQ-TREE (Nguyen et al., 2015) and visualized with ITOL (Letunic and Bork, 2019). The multiple alignment consisted of 253 sequences and 774 amino acid sites, aligned using MAFFT and trimmed with Trimal (Capella-Gutiérrez et al., 2009; Katoh and Standley, 2013). The best fit model was FLU+F+R8 as determined with ModelFinder (Kalyaanamoorthy et al., 2017). Branch support was calculated using the UFBoot (ultrafast bootstrap) algorithm on 1000 bootstraps and is indicated with branch colours in shades of grey, with support values higher than 95% in black (Hoang et al., 2018). Colour strip 1 indicates the genotype clustering, using RVC isolates as outgroup. Colour strip 2 shows the host of the isolates with arrows indicating the virome-derived sequences.

infants and children under the age of five (Hamborsky et al., 2015), who are less likely to engage with such activities which may be the reason for the lack of reported illnesses.

We identified a number of small contigs related to ssDNA circular circoviruses and parvoviruses, that were originally recovered from environmental or host-associated metagenomes (Dayaram et al., 2015; Phan et al., 2015; Zawar-Reza et al., 2014). Of these, four circovirus-associated vOTUs, representing 18 contigs, showed significant sequence similarity to previously described UViGs from animal or wastewater metagenomes. One parvovirus contig, assigned to the genus *Ambidensovirus*, was related to a bat metagenome sequence. However, for these types of ssDNA virus UViGs, any causative links with disease syndromes would be very tenuous.

3.3. *A conceptual model for virus circulation in a freshwater catchment area*

The data presented in this study support the following conceptual model of virus circulation in the river system (Fig. 6). Upstream, in the more pristine regions of the river with low human and livestock inputs, viral species richness is low and the water virome is dominated by dsDNA tailed bacteriophages (caudoviruses) and a few algal viruses of

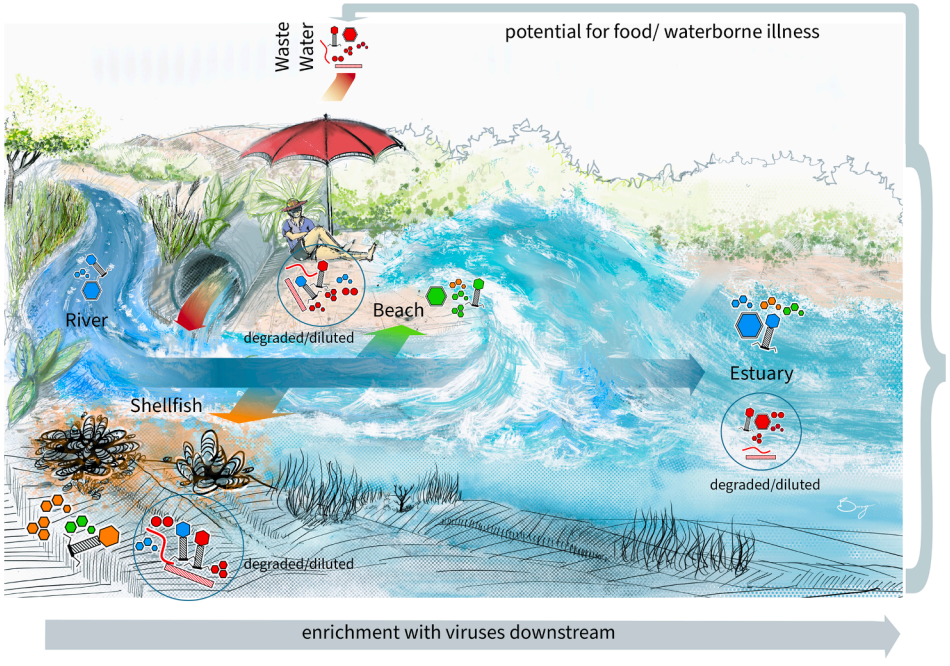

Fig. 6. Model for the circulation of viruses in a river catchment and coastal zone system with wastewater discharge. Viruses specific to river water are depicted in blue, wastewater in red, beach sediment in green and shellfish in orange.

the family *Phycodnaviridae*. At certain points along the river, wastewater effluent from large treatment plants and smaller scale septic tank discharges enter the water. This effluent is much less rich in viruses than untreated wastewater (influent) but can still contain over 1000 different viral species per litre. The entire spectrum of viral diversity detected in this study is represented in effluent (treated wastewater), with DNA and RNA bacteriophages (predicted to infect members of the human gut microbiome) the most commonly detected groups (caudoviruses, leviviruses). Nucleo-cytoplasmic large DNA virus (NCLDVs; phycodnaviruses, mimiviruses, iridoviruses) and common plant-derived viruses present in food and excreted by the human digestive tract (mainly tobamoviruses such as pepper mild mottle virus (Rosario et al., 2009; Zhang et al., 2006)) and groups of enteric viruses such as sapovirus, rotavirus and astrovirus within a wider collection of unclassified RNA viruses are also well represented. These communities are both spatially and temporally distinct. Upon entering the river, the pathogenic virus groups fall below the limit of detection by virome sequencing, which can be attributed primarily to dilution by the river water. However, close to an effluent site and at the estuary that is under tidal control, the number of viral species detected in water samples is much higher. Beach sediment and filter-feeding shellfish (in this case mussels, *Mytilus edulis*) then act as entrapment matrices enriching the viral content from the surrounding water (Maalouf et al., 2010; Whitman et al., 2014). In the majority of cases, the UViGs that were assembled from wastewater recruited fewer reads from beach sediment, mussel tissue or estuary water libraries, and the read mapping over the genome length was often patchy, leading us to hypothesize that these genomes, and by extension the virions, are likely to be substantially degraded. At the same time, we observed sediment- and mussel-specific viral communities represented by full genomes, mainly picorna-like RNA viruses and unclassified UViGs from invertebrates (Shi et al., 2016), thus excluding technical bias as the explanantion for our failure to detect intact pathogenic virus genomes in sediment and shellfish. In the scenario that we propose, shellfish and sediment become enriched in viruses that are recruited from the environment by filter feeding and adsorption, respectively. Those viruses that do not undergo active replication in the newly occupied niche (human, animal and plant pathogens in particular) are degraded over time or diluted below the limit of detection, while viruses that infect the shellfish, the shellfish microbiome, diatoms or sediment-associated bacteria are maintained, enabling detection of their full genome sequences. In this scenario, the risk of illness due to consumption of shellfish, contact with sediment (beach sand) or swimming, would depend on the time interval between uptake/adsorption of pathogenic viruses in the matrix and ingestion by a human subject. To critically evaluate this, further experimental data on the infectivity/-survival kinetics for each viral species are required, as this is likely to vary markedly between viral groups. However, this would be a Herculean endeavour, given the diversity of viruses detected here, the difficulty in propagation and the absence of routine infectivity assays. The conceptual model is supported by the results of our previous year-long q (RT)-PCR study on a subset of enteric viruses, which showed that they were still detected at high titres in wastewater post-treatment, followed by lower titres in river water, shellfish and sediment, and ultimately undergoing capsid degradation in environmental matrices (Farkas et al., 2018a). Our conceptual model of viral circulation is also consistent with theorectical simulations of viral discharge from wastewater treatment plants into the coastal zone (Robins et al., 2019). Importantly, these models have indicated that tidal movement allows viruses in estuarine water to come into contact with shellfisheries and beaches on numerous occasions over a period of days to weeks depending on the lunar tidal cycle.

4. Conclusion

Viruses and their genetic material are commonly discharged in the environment, but their risk to human health is driven by community outbreaks leading to viral shedding into the wastewater, leading to temporal and spatial variations in the specific genotypes detected. In the environment, these viruses are then subject to cycles of dilution, enrichment and virion degradation influenced by local geography, weather events and tidal effects. Our analyses show that viromics is a useful tool to assess viral diversity in the aquatic environment in order to explore new and emerging human and animal health threats.

Availability of data and material

The datasets generated and analysed in this study are available from the Sequence Read Archive (SRA) under BioProject PRJNA509142, accession numbers SRR8299359 to SRR8299398.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dwr Cymru/Welsh Water for access to the wastewater treatment plants. We thank the following people for assistance with sample collection: Dr Emma Green and Harry Riley, Bangor University. This work was supported by the Natural Environment Research Council (NERC) and the Food Standards Agency (FSA) under the Environmental Microbiology and Human Health (EMHH) Programme (VIRAQUA; NE/M010996/1). E.M.A is currently funded by the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Gut Microbes and Health (BB/R012490/1) and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2021.117568.

References

Adriaenssens, E., Brister, J.R., 2017. How to name and classify your phage: An informal guide. Viruses 9, 70. https://doi.org/10.3390/v9040070.

Adriaenssens, E.M., Farkas, K., Harrison, C., Jones, D.L., Allison, H.E., McCarthy, A.J., 2018. Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses. mSystems 3. https://doi.org/10.1128/mSystems.00025-18 e00025-18.

Ahmed, S.M., Hall, A.J., Robinson, A.E., Verhoef, L., Premkumar, P., Parashar, U.D., Koopmans, M., Lopman, B.A., 2014. Global prevalence of norovirus in cases of gastroenteritis: A systematic review and meta-analysis. Lancet Infect. Dis. 14, 725–730. https://doi.org/10.1016/S1473-3099(14)70767-4.

Amar, C.F.L., East, C.L., Gray, J., Iturriza-Gomara, M., Maclure, E.A., McLauchlin, J., 2007. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: Re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). Eur. J. Clin. Microbiol. Infect. Dis. 26, 311–323. https://doi.org/10.1007/s10096-007-0290-8.

Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699. https://doi.org/10.1093/sysbio/syr041.

Anisomova, M., Gascuel, O., 2006. Approximate Likelihood-Ratio Test for Branches : A Fast, Accurate,. Syst. Biol. 55, 539–552. https://doi.org/10.1080/10635150600755453.

Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B.L., Sullivan, M.B., 2017. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. ISME J 11, 7–14. https://doi.org/10.1038/ismej.2016.89.

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519.

Brown, J.R., Shah, D., Breuer, J., 2016. Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015. J. Clin. Virol. 84, 1–6. https://doi.org/10.1016/j.jcv.2016.08.298.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Da Silva, A.K., Le Saux, J.C., Parnaudeau, S., Pommepuy, M., Elimelech, M., Le Guyader, F.S., 2007. Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: Different behaviors of genogroups I and II. Appl. Environ. Microbiol. 73, 7891–7897. https://doi.org/10.1128/AEM.01428-07.

Dayaram, A., Goldstien, S., Argüello-Astorga, G.R., Zawar-Reza, P., Gomez, C., Harding, J.S., Varsani, A., 2015. Diverse small circular DNA viruses circulating amongst estuarine molluscs. Infect. Genet. Evol. 31, 284–295. https://doi.org/10.1016/j.meegid.2015.02.010.

DiCaprio, E., 2017. Recent advances in human norovirus detection and cultivation methods. Curr. Opin. Food Sci. 14, 93–97. https://doi.org/10.1016/j.cofs.2017.02.007.

Diez-Valcarce, M., Castro, C.J., Marine, R.L., Halasa, N., Mayta, H., Saito, M., Tsaknaridis, L., Pan, C.Y., Bucardo, F., Becker-Dreps, S., Lopez, M.R., Magaña, L.C., Ng, T.F.F., Vinjé, J., 2018. Genetic diversity of human sapovirus across the Americas. J. Clin. Virol. 104, 65–72. https://doi.org/10.1016/j.jcv.2018.05.003.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.., Jang, H.B., Singleton, C. M., Solden, L.M., Naas, A.E., Boyd, J.A., Hodgkins, S.B., Wilson, R.M., Trubl, G., Li, C., Frolking, S., Pope, P.B., Wrighton, K.C., Crill, P.M., Chanton, J.P., Sullivan, M. B., 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nat. Microbiol. https://doi.org/10.1038/s41564-018-0190-y.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T. O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319. https://doi.org/10.7717/peerj.1319.

Farkas, K., Adriaenssens, E.M., Walker, D.I., McDonald, J.E., Malham, S.K., Jones, D.L., 2019. Critical Evaluation of CrAssphage as a Molecular Marker for Human-Derived Wastewater Contamination in the Aquatic Environment. Food Environ. Virol. 0 https://doi.org/10.1007/s12560-019-09369-1, 0.

Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A., Jones, D.L., 2018a. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Sci. Total Environ. 634, 1174–1183. https://doi.org/10.1016/j.scitotenv.2018.04.038.

Farkas, K., Hassard, F., McDonald, J.E., Malham, S.K., Jones, D.L., 2017a. Evaluation of molecular methods for the detection and quantification of pathogen-derived nucleic acids in sediment. Front. Microbiol. 8, 53. https://doi.org/10.3389/fmicb.2017.00053.

Farkas, K., Mannion, F., Hillary, L.S., Malham, S.K., Walker, D.I., 2020. Emerging technologies for the rapid detection of enteric viruses in the aquatic environment. Curr. Opin. Environ. Sci. Heal. https://doi.org/10.1016/j.coesh.2020.01.007.

Farkas, K., Marshall, M., Cooper, D., McDonald, J.E., Malham, S.K., Peters, D.E., Maloney, J.D., Jones, D.L., 2018b. Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. Environ. Sci. Pollut. Res. 33391–33401. https://doi.org/10.1007/s11356-018-3261-y.

Farkas, K., McDonald, J.E., Malham, S.K., Jones, D.L., 2018c. Two-step concentration of complex water samples for the detection of viruses. Methods Protoc 1, 35.

Farkas, K., Peters, D.E., McDonald, J.E., de Rougemont, A., Malham, S.K., Jones, D.L., 2017b. Evaluation of Two Triplex One-Step qRT-PCR Assays for the Quantification of Human Enteric Viruses in Environmental Samples. Food Environ. Virol. 9, 342–349. https://doi.org/10.1007/s12560-017-9293-5.

Finkbeiner, S.R., Kirkwood, C.D., Wang, D., 2008. Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. Virol. J. 5, 117. https://doi.org/10.1186/1743-422X-5-117.

Fong, T.T., Phanikumar, M.S., Xagoraraki, I., Rose, J.B., 2010. Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a Michigan river. Appl. Environ. Microbiol. 76, 715–723. https://doi.org/10.1128/AEM.01316-09.

FSA, 2017. Estimating Quality Adjusted Life Years and Willingness to Pay Values for Microbiological Foodborne Disease (Phase 2). London, UK.

Girones, R., Ferrús, M.A., Alonso, J.L., Rodriguez-Manzano, J., Calgua, B., de Abreu Corrêa, A., Hundesa, A., Carratala, A., Bofill-Mas, S., 2010. Molecular detection of pathogens in water - The pros and cons of molecular techniques. Water Res 44, 4325–4339. https://doi.org/10.1016/j.watres.2010.06.030.

Gomes, J., Frasson, D., Quinta-Ferreira, R., Matos, A., Martins, R., 2019. Removal of Enteric Pathogens from Real Wastewater Using Single and Catalytic Ozonation. Water 11, 127. https://doi.org/10.3390/w11010127.

Gorbalenya, A.E., Krupovic, M., Mushegian, A., Kropinski, A.M., Siddell, S.G., Varsani, A., Adams, M.J., Davison, A.J., Dutilh, B.E., Harrach, B., Harrison, R.L., Junglen, S., King, A.M.Q., Knowles, N.J., Lefkowitz, E.J., Nibert, M.L., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., Walker, P.J., Zerbini, F.M., Kuhn, J. H., 2020. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat. Microbiol. 5, 668–674. https://doi.org/10.1038/s41564-020-0709-x.

Gregory, A.C., Solonenko, S.A., Ignacio-Espinoza, J.C., LaButti, K., Copeland, A., Sudek, S., Maitland, A., Chittick, L., dos Santos, F., Weitz, J.S., Worden, A.Z., Woyke, T., Sullivan, M.B., 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. BMC Genomics 17, 930. https://doi.org/10.1186/s12864-016-3286-x.

Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., Sullivan, M.B., 2020. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host Microbe 28, 724–740. https://doi.org/10.1016/j.chom.2020.08.003 e8.

Gulino, K., Rahman, J., Badri, M., Morton, J., Bonneau, R., Ghedin, E., 2020. Initial Mapping of the New York City Wastewater Virome. mSystems 5, 1–18. https://doi.org/10.1128/msystems.00876-19.

Hamborsky, J., Kroger, A., Wolfe, C, 2015. Epidemiology and prevention of vaccine-preventable diseases, 13th Editi. Centers for Disease Control and Prevention, Washington DC, USA.

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl. Environ. Microbiol. 80, 6771–6781. https://doi.org/10.1128/AEM.01981-14.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Huson, D.H., Weber, N., 2013. Microbial Community Analysis Using MEGAN, in: Methods in Enzymology. pp. 465–485. https://doi.org/10.1016/B978-0-12-407863-5.00021-6.

Inns, T., Wilson, D., Manley, P., Harris, J.P., O'Brien, S.J., Vivancos, R., 2019. What proportion of care home outbreaks are caused by norovirus? An analysis of viral causes of gastroenteritis outbreaks in care homes, North East England, 2016-2018. BMC Infect. Dis. 20, 1–8. https://doi.org/10.1186/s12879-019-4726-4.

Joshi, N., Fass, J., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software].

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleesschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C.L., Hald, T., Hall, A.J., Keddy, K.H., Lake, R. J., Lanata, C.F., Torgerson, P.R., Havelaar, A.H., Angulo, F.J., 2015. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med 12, 1–21. https://doi.org/10.1371/journal.pmed.1001921.

Kitajima, M., Iker, B.C., Pepper, I.L., Gerba, C.P., 2014. Relative abundance and treatment reduction of viruses during wastewater treatment processes—Identification of potential viral indicators. Sci. Total Environ. 488, 290–296. https://doi.org/10.1016/j.scitotenv.2014.04.087.

Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., Kuhn, J.H., 2020. Global Organization and Proposed Megataxonomy of the Virus World. Microbiol. Mol. Biol. Rev. 84 https://doi.org/10.1128/MMBR.00061-19 e00061-19.

Kroger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hebrard, M., Handler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G.C., Lohan, A.J., Loftus, B., Lucchini, S., Ussery, D.W., Dorman, C.J., Thomson, N.R., Vogel, J., Hinton, J.C.D., 2012. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proc. Natl. Acad. Sci. 109, E1277–E1286. https://doi.org/10.1073/pnas.1201061109.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47, W256–W259. https://doi.org/10.1093/nar/gkz239.

Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128. https://doi.org/10.1093/bioinformatics/btl529.

Maalouf, H., Zakhour, M., Pendu, J.Le, Le Saux, J.C., Atmar, R.L., Le Guyader, F.S., 2010. Distribution in tissue and seasonal variation of norovirus genogroup I and II ligands in oysters. Appl. Environ. Microbiol. 76, 5621–5630. https://doi.org/10.1128/AEM.00148-10.

Maes, P., Matthijnssens, J., Rahman, M., Ranst, M. Van, 2009. RotaC : A web-based tool for the complete genome classification of group A rotaviruses 4, 2–5. https://doi.org/10.1186/1471-2180-9-238.

Mann, Pietsch, Liebert, 2019. Genetic Diversity of Sapoviruses among Inpatients in Germany, 2008−2018. Viruses 11, 726. https://doi.org/10.3390/v11080726.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200.

Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7, 1–9. https://doi.org/10.1038/ncomms11257.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Noel, J.S., Liu, B.L., Humphrey, C.D., Rodriguez, E.M., Lambden, P.R., Clarke, I.N., Dwyer, D.M., Ando, T., Glass, R.I., Monroe, S.S., 1997. Parkville virus: A novel genetic variant of human calicivirus in the Sapporo virus clade, associated with an outbreak of gastroenteritis in adults. J. Med. Virol. 52, 173–178. https://doi.org/10.1002/(SICI)1096-9071(199706)52:2<173::AID-JMV10>3.0.CO;2-M.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel, S.R., Woyke, T., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. In: Deng, M., Jiang, R., Sun, F., Zhang, X. (Eds.), Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 158–170. https://doi.org/10.1007/978-3-642-37195-0_13.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O.G., Kandrov, D., Rasputin, K., Syabro, M., Tleukenov, T., 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28, 1166–1167. https://doi.org/10.1093/bioinformatics/bts091.

Pang, X., Qiu, Y., Gao, T., Zurawell, R., Neumann, N.F., Craik, S., Lee, B.E., 2019. Prevalence, levels and seasonal variations of human enteric viruses in six major rivers in Alberta, Canada. Water Res. 153, 349–356. https://doi.org/10.1016/j.watres.2019.01.034.

Pérez-Cataluña, A., Cuevas-Ferrando, E., Randazzo, W., Sánchez, G., 2021. Bias of library preparation for virome characterization in untreated and treated wastewaters. Sci. Total Environ. 767, 144589 https://doi.org/10.1016/j.scitotenv.2020.144589.

Perkins, T.L., Clements, K., Baas, J.H., Jago, C.F., Jones, D.L., Malham, S.K., McDonald, J.E., 2014. Sediment Composition Influences Spatial Variation in the Abundance of Human Pathogen Indicator Bacteria within an Estuarine Environment. PLoS One 9, e112951. https://doi.org/10.1371/journal.pone.0112951.

Phan, T.G., Mori, D., Deng, X., Rajidrajith, S., Ranawaka, U., Fan Ng, T.F., Bucardo-Rivera, F., Orlandi, P., Ahmed, K., Delwart, E., 2015. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. Virology 482, 98–104. https://doi.org/10.1016/j.virol.2015.03.011.

Prado, T., de Castro Bruni, A., Barbosa, M.R.F., Garcia, S.C., de Jesus Melo, A.M., Sato, M.I.Z., 2019. Performance of wastewater reclamation systems in enteric virus removal. Sci. Total Environ. 678, 33–42. https://doi.org/10.1016/j.scitotenv.2019.04.435.

Qiu, Y., Lee, B.E., Neumann, N., Ashbolt, N., Craik, S., Maal-Bared, R., Pang, X.L., 2015. Assessment of human virus removal during municipal wastewater treatment in Edmonton, Canada. J. Appl. Microbiol. 119, 1729–1739. https://doi.org/10.1111/jam.12971.

Robins, P.E., Farkas, K., Cooper, D., Malham, S.K., Jones, D.L., 2019. Viral dispersal in the coastal zone: A method to quantify water quality risk. Environ. Int. 126, 430–442. https://doi.org/10.1016/j.envint.2019.02.042.

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.

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., Amid, C., Aziz, R.K., Bordenstein, S.R., Bork, P., Breitbart, M., Cochrane, G.R., Daly, R.A., Desnues, C., Duhaime, M.B., Emerson, J.B., Enault, F., Fuhrman, J.A., Hingamp, P., Hugenholtz, P., Hurwitz, B.L., Ivanova, N.N., Labonté, J.M., Lee, K.-B., Malmstrom, R.R., Martinez-Garcia, M., Mizrachi, I.K., Ogata, H., Páez-Espino, D., Petit, M.-A., Putonti, C., Rattei, T., Reyes, A., Rodriguez-Valera, F., Rosario, K., Schriml, L., Schulz, F., Steward, G.F., Sullivan, M.B., Sunagawa, S., Suttle, C.A., Temperton, B., Tringe, S.G., Thurber, R.V., Webster, N.S., Whiteson, K.L., Wilhelm, S.W., Wommack, K.E., Woyke, T., Wrighton, K.C., Yilmaz, P., Yoshida, T., Young, M.J., Yutin, N., Allen, L.Z., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat. Biotechnol. 37, 29–37. https://doi.org/10.1038/nbt.4306.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J., Pesant, S., Kandels-Lewis, S., Dimier, C., Picheral, M., Searson, S., Cruaud, C., Alberti, A., Duarte, C.M., Gasol, J.M., Vaqué, D., Bork, P., Acinas, S.G., Wincker, P., Sullivan, M.B., 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537, 689–693. https://doi.org/10.1038/nature19366.

Roux, S., Enault, F., Hurwitz, B.L., Sullivan, M.B., 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3, e985. https://doi.org/10.7717/peerj.985.

Roux, S., Páez-Espino, D., Chen, I.A., Palaniappan, K., Ratner, A., Chu, K., Reddy, T.B.K., Nayfach, S., Schulz, F., Call, L., Neches, R.Y., Woyke, T., Ivanova, N.N., Eloe-Fadrosh, E.A., Kyrpides, N.C., 2020. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Res 1–12. https://doi.org/10.1093/nar/gkaa946.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864. https://doi.org/10.1093/bioinformatics/btr026.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., Holmes, E.C., Zhang, Y.-Z., 2018. The evolutionary history of vertebrate RNA viruses. Nature 556, 197–202. https://doi.org/10.1038/s41586-018-0012-7.

Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.-Z., 2016. Redefining the invertebrate RNA virosphere. Nature 540, 1–12. https://doi.org/10.1038/nature20167.

Sidhu, J.P.S., Sena, K., Hodgers, L., Palmer, A., Toze, S., 2017. Comparative enteric viruses and coliphage removal during wastewater treatment processes in a sub-tropical environment. Sci. Total Environ. 616, 669–677. https://doi.org/10.1016/j.scitotenv.2017.10.265.

Soneson, C., Love, M.I., Robinson, M.D., 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4, 1521. https://doi.org/10.12688/f1000research.7563.2.

Tapparel, C., Siegrist, F., Petty, T.J., Kaiser, L., 2013. Picornavirus and enterovirus diversity with associated human diseases. Infect. Genet. Evol. 14, 282–293. https://doi.org/10.1159/000445409.

Varela, M.F., Ouardani, I., Kato, T., Kadoya, S., Aouni, M., Sano, D., Romalde, J.L., 2018. Sapovirus in wastewater treatment plants in Tunisia: Prevalence, removal, and genetic characterization. Appl. Environ. Microbiol. 84 https://doi.org/10.1128/AEM.02093-17.

Venkataraman, S., Prasad, B., Selvarajan, R., 2018. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. Viruses 10, 76. https://doi.org/10.3390/v10020076.

Vinjé, J., Deijl, H., Van Der Heide, R., Lewis, D., Hedlund, K.O., Svensson, L., Koopmans, M.P.G., 2000. Molecular detection and epidemiology of Sapporo-like viruses. J. Clin. Microbiol. 38, 530–536. https://doi.org/10.1128/jcm.38.2.530-536.2000.

Whitman, R.L., Harwood, V.J., Edge, T.A., Nevers, M.B., Byappanahalli, M., Vijayavel, K., Brandão, J., Sadowsky, M.J., Alm, E.W., Crowe, A., Ferguson, D., Ge, Z., Halliday, E., Kinzelman, J., Kleinheinz, G., Przybyla-Kelly, K., Staley, C., Staley, Z., Solo-Gabriele, H.M., 2014. Microbes in beach sands: Integrating environment, ecology and public health, Reviews in Environmental Science and Biotechnology. https://doi.org/10.1007/s11157-014-9340-8.

Wickham, H., 2016. GGPLOT2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, NY, USA.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686.

Zawar-Reza, P., Argüello-Astorga, G.R., Kraberger, S., Julian, L., Stainton, D., Broady, P.A., Varsani, A., 2014. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2014.05.018.

Zell, R., Delwart, E., Gorbalenya, A.E., Hovi, T., King, A.M.Q., Knowles, N.J., Lindberg, A.M., Pallansch, M.A., Palmenberg, A.C., Reuter, G., Simmonds, P., Skern, T., Stanway, G., Yamashita, T., 2017. ICTV Virus Taxonomy Profile: Picornaviridae. J. Gen. Virol. 98, 2421–2422. https://doi.org/10.1099/jgv.0.000911.

Zenkova, D., Kamenev, V., Sablina, R., Artyomov, M., Sergushichev, A., 2018. Phantasus: visual and interactive gene expression analysis. https://doi.org/10.18129/B9.bioc.phantasus.

Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2006. RNA viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol 4, 0108–0118. https://doi.org/10.1371/journal.pbio.0040003.