Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance

Authorship

X. Fernandez-Cassi[1,*,¶] N. Timoneda[1,2,¶], S. Martínez-Puchol[1], M. Rusiñol[1], J. Rodriguez-Manzano[3], N. Figuerola[1], S. Bofill-Mas[1], J.F. Abril[2], R. Girones[1].

[1] Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain.

[2]Computational Genomics Lab, University of Barcelona and Institute of Biomedicine (IBUB), University of Barcelona; Barcelona, Catalonia, Spain.

[3]Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom.

[*] Corresponding author.

[¶] These authors contributed equally to this work.

Abstract/summary

The application of next-generation sequencing (NGS) techniques for the identification of viruses present in urban sewage has not been fully explored. This is partially due to a lack of reliable and sensitive protocols for studying viral diversity and to the highly complex analysis required for NGS data processing. One important step towards this goal is finding methods that can efficiently concentrate viruses from sewage samples. Here the application of a virus concentration method based on skimmed milk organic flocculation (SMF) using 10 L of sewage collected in different seasons enabled the detection of many viruses. However, some viruses, such as human adenoviruses, could not always be detected using metagenomics , even when qPCR assessments were positive. A targeted metagenomic assay for adenoviruses was conducted and 59.41% of the obtained reads were assigned to murine adenoviruses. However, up to 20 different human adenoviruses (HAdV) were detected by this targeted assay being the most abundant HAdV-41 (29.24 %) and HAdV-51 (1.63%). To improve metagenomics' sensitivity, two different protocols for virus concentration were comparatively analysed: an ultracentrifugation protocol and a lower-volume SMF protocol. The sewage virome contained 41 viral families, including pathogenic viral species from families *Caliciviridae, Adenoviridae, Astroviridae, Picornaviridae, Polyomaviridae, Papillomaviridae* and *Hepeviridae*. The contribution of urine to sewage metavirome seems to be restricted to a few specific DNA viral

families, including the polyomavirus and papillomavirus species. In experimental infections with sewage in a rhesus macaque model, infective human hepatitis E and JC polyomavirus were identified. Urban raw sewage consists of the excreta of thousands of inhabitants; therefore, it is a representative sample for epidemiological surveillance purposes. The knowledge of the metavirome is of significance to public health, highlighting the presence of viral strains that are circulating within a population while acting as a complex matrix for viral discovery.

Keywords: viral metagenomics, human adenovirus, viral pathogens, sewage, next-generation sequencing

1. Introduction

In recent years, water scarcity and the application of more sustainable water reuse practices has favoured the utilisation of treated sewage for several purposes, including crop and green area irrigation, river catchment replenishment and toilet flushing. Conventional treatments applied in sewage treatment plants (STPs) are known to be less efficient for virus removal compared to faecal indicator bacteria (FIB) (Gerba et al., 1979; Pusch et al., 2005). This higher viral survival in STP treated water can represent a threat to consumers because STP effluents which contain viruses can contaminate water and food. Raw urban sewage is a complex matrix consisting of urine, faeces and skin desquamation from people. Therefore, raw sewage contains a large variety of pathogenic and commensal viruses, bacteria and protozoa excreted from thousands of inhabitants. Additionally, a high number of plant viruses pass though the human intestines. Sewage also contains other non-human inputs, which increase the diversity of this complex ecosystem. Viruses do not have a conserved molecular markers, such as 16S rRNA, that are shared across all species and hampers the study of viral metagenomes. However, the application of random-primer-based sequencing approaches in combination with next-generation sequencing (NGS) techniques has opened a new path for viral discovery, increasing number of new viral species described each year. Viral metagenomics applications to sewage provide excellent tools for monitoring and identifying potentially known and unknown viral pathogens that circulate among the human population, contributing to public health surveillance.

Although some viral metagenomics protocols are available for clinical samples (Kohl et al., 2015), only a few manuscripts describe the application of metagenomics approaches to analyse the viruses present in sewage (Cantalupo et al., 2011; Ng et al., 2012). Previous studies have shown that viruses prevalent in sewage are not always detected by metagenomics, suggesting that protocols should be improved to increase sensitivity. For example, in a study by Cantalupo and collaborators, showed that human adenoviruses (HAdVs) in sewage samples were barely detectable by NGS while by quantitative PCR (qPCR) the HAdV genome was highly prevalent.

In this study, we investigated the diversity of viruses present in raw sewage using metagenomics to test samples from three different seasons. The application of this methodology allowed for a description of the human virome and an evaluation of the sensitivity of the technique using HAdVs as a reference. HAdVs were selected because of their dual role as pathogens and specific human viral faecal indicators (Bofill-Mas et al., 2013).

With this purpose, we compared the performance of an untargeted metagenomics analysis to an adenovirus-targeted NGS assay and HAdV qPCR values. To increase the number of different viral species identified in sewage, different protocols for concentrating viruses in urban sewage were evaluated, and an efficient protocol for the analysis of viruses in sewage and other environmental samples by metagenomics has been proposed.

The application of metagenomics in different human body parts has facilitated the study of viral communities in the oral cavity (Ly et al., 2014), gut (Minot et al., 2011), respiratory tract (Willner et al., 2009), skin (Foulongne et al., 2012), blood (Sauvage et al., 2016) and cerebrospinal fluid (Perlejewski et al., 2015). Viral faecal viromes have been studied in healthy (Minot et al., 2011) and unhealthy patients (Linsuwanon et al., 2015) and in domestic animals (Mihalov-Kovács et al., 2014). Hence, the viral contribution of faeces to raw sewage seems clear. Of note, the viral communities excreted through urine remain poorly studied (Santiago-Rodriguez et al., 2015), which may be because urine has typically been considered a sterile environment. To assess the contribution of urine to the virome of raw sewage and to study the viral composition of the urine, viruses in pooled urine samples were also analysed by metagenomics in this study.

The infectivity of the known and unknown viral species present in raw sewage was explored by intravenously inoculating a sewage sample into rhesus monkeys as a potential enrichment step prior to the use of the metagenomics to examine rhesus serum samples.

Finally, a tailored protocol to analyse sewage and other environmental samples using metagenomics was proposed. Bioinformatics-specific parameters were adjusted to different levels, and new tools were tested to filter out the best set of raw reads, such as those containing the most informative sequences. These reads were combined into assembled contigs that were later used to detect the known species genomes present in the samples and relative abundances of the taxonomic groups found in the species mixture. Similarity searches also provided a basic characterisation of the pathogenic species present in the samples.

2. Materials and methods

2.1 Concentration of viral particles from tested samples

2.1.1. Concentration of viral particles from raw sewage using skimmed milk organic flocculation.

Sampling points of this study are presented in Figure 1. Three 10-L samples of raw sewage from a STP in Sant Adrià del Besós were collected in winter, spring and summer of 2013. Samples were processed 2 h after collection and stored at -80°C until further analysis. Viral particles were concentrated using the SMF method described by Cantalupo et al. (2011). Free DNA from

viral concentrates was removed, nucleic acids (NAs) were extracted, and libraries were prepared as explained in section 2.2.

In the second protocol, a reduction in the sewage sample volume was also evaluated in order to reduce the levels of inhibitory compounds and interfering materials in the viral concentrate.

The SMF-adapted protocol used 500 mL of raw sewage that was preconditioned to a pH 3.5 and was based on the protocol described by Calgua et al., 2008. Briefly, 5 mL of a pre-flocculated skim milk solution at pH 3.5 and a conductivity superior to 1.5 mS/cm^2 was added to each sample. After 8 h of stirring, flocks were centrifuged at 8000xg for 40 minutes, and the pellet was suspended in 4 mL of phosphate buffer [vol/vol] (0.2 M Na_2HPO_4 and 0.2 M NaH_2PO_4). The viral concentrate was kept at -80°C until further use.

A third protocol based on ultracentrifugation was evaluated in comparison to the 500 mL SMF protocol. Two 600-mL samples of raw sewage were collected from Granollers STP. Samples were divided into two aliquots: 500 mL for processing according to the SMF protocol adapted from Calgua et al., 2008 and 42 mL for the ultracentrifugation protocol adapted from Pina et al., 1998a. The obtained SMF and ultracentrifugation viral concentrates were filtered through 0.45-µm Sterivex filters (Millipore, Massachusetts). Free DNA was removed, NAs were extracted, and libraries were prepared as explained in section 2.2. For both methodologies, the equivalent of 7 mL of a raw sewage sample was analysed in the final constructed libraries. The presence of HAdV was analysed by qPCR on the NA extracts as described by Bofill-Mas et al., 2006.

2.1.2 Concentration of viral particles from urine

To explore the viruses excreted in urine and the urine contribution to the raw sewage virome, 100 mL of urine was collected from 14 healthy volunteers of various ages and origins (7 males and 7 females from 25-63 years old), although most of them were living in Barcelona. Individual urine samples were ultracentrifuged for 1 h at 90,000xg at 4°C. The obtained viral pellets were suspended in 300 µL of 1X PBS and kept at -80°C until further use. A pooled sample with 1000 µL of urine viral concentrate was obtained by pooling individual samples already processed. From the pooled sample, 500 µL was treated with DNase, NAs were extracted, and a library was prepared as explained in section 2.2.

2.2 Free DNA removal, nucleic acid extraction, library preparation and sequencing

For all samples, DNase treatments were performed using the same conditions. Then, 300 µL of raw sewage viral concentrate was treated with 160 U of Turbo DNase (Ambion Cat no. AM1907, Ambion) for 1 h at 37°C to remove non-viral free DNA. DNase was inactivated using the provided inactivation reagent, and the samples were centrifuged at 10,000xg for 1.5 minutes. The treated supernatants were collected and kept at 4°C until the nucleic viral acid extraction. Then, 280 µL of the viral concentrate was extracted using the Qiagen RNA Viral Mini Kit (cat no. 22906, Qiagen, Valencia, CA, USA) without the RNA carrier. NAs were eluted using 60 µL of AVE buffer.

For all samples, libraries were prepared following the same protocol. To detect both RNA and DNA viruses, NAs were retrotranscribed using random nonamer Primer A (5'-GTTTCCCAGTCACGATANNNNNNNNN'-3) as previously described in Wang et al., 2003. Briefly, RNA templates were reverse transcribed using SuperScript III (cat no. 18080093, Life Technologies) and Primer A, which contains a 17-nucleotide-specific sequence followed by 9 random nucleotides for random priming. A second cDNA strand was constructed using Sequenase 2.0 (cat no. USBM70775Y200UN, USB/Affymetrix, Cleveland, OH, USA). To obtain sufficient DNA for library preparation, a PCR amplification step using Primer B (5'-GTTTCCCAGTCACGATANNNNNNNNN'-3) and AmpliTaqGold (cat no. 4311806, Life Technologies, Austin, Texas, USA) was performed. After 10 min at 95°C to activate DNA polymerase, the following PCR programme was used: 25 cycles of 30 s at 94°C, 30 s at 40°C, and 30 s at 50°C for the ultracentrifugation and low-volume adapted SMF and 40 cycles of 30 s at 94°C, 30 s at 40°C, and 30 s at 50°C for the 10-L SMF protocol, with a final step of 60 s at 72°C for all protocols. PCR products were cleaned and concentrated in a small volume (15 µL) using the Zymo DNA clean and concentrator (D4013, Zymo research, USA). Amplified DNA samples were quantified using Qubit 2.0 (cat no. Q32854, Life Technologies, Oregon, USA), and libraries were constructed using a Nextera XT DNA sample preparation kit (Illumina Inc) according to the manufacturer's instructions. Samples were sequenced on Illumina MiSeq 2x250 bp and 2x300 bp, producing paired end reads.

2.3. Bioinformatic pipeline and quality filtering

The quality of raw and clean read sequences was assessed using the FASTX-Toolkit software, version 0.0.14 (Hannon Lab, http://www.hannonlab.org). Read sequences were cleaned using Trimmomatic, version 0.32 (Bolger et al., 2014), taking care of sequencing adaptors and linker

contamination. Low quality ends were trimmed per an average threshold Phred score above Q15 over a running-window of 4 nucleotides. Low complexity sequences, which were mostly biased to repetitive sequences that affect the performance of downstream computational procedures, were then discarded after estimating a linear model based on Trifonov's linguistic complexity (Sarma et al., 1990) and the sequence string compression ratio. Discrimination criteria for the linear model assumed low complexity scores below a line with a 45° slope and crossing at 5% below the complexity inflexion point found by the model, which was specific to each sequence set. Finally, duplicated reads were removed in a subsequent step to accelerate the downstream assembly. Virome reads were assembled based on 90% identity over a minimum of 50% of the read length using CLC Genomics Workbench 4.4 (CLC bio USA, Cambridge, MA), and the resulting contig spectra were used as the primary input for the index. Afterwards, contigs longer than 100 bp were queried for sequence similarity using BLASTN and BLASTX (Altschul et al., 1997, 1990) against the NCBI viral complete genomes database (Brister et al., 2015), the viral division from GenBank nucleotide database (Benson et al., 2015), and the viral protein sequences from UniProt (UniProt Consortium 2015, ftp://ftp.uniprot.org/pub/databases/uniprot/current_release). The species nomenclature and classification were performed according to the NCBI Taxonomy database standards. HSPs considered for taxonomical assessment had an E-value of 10^{-5} and minimum length of 100 bp. Based on the best BLAST result and 90% coverage cut-off, each sequence was classified into its likely taxonomic group-of-origin. Tables summarizing the number of sequences from the assembly matching each taxonomic unit were built. From these tables, richness ratios were calculated using the Catchall software, version 4.0 (Allen et al., 2013). Of the models included in the package, the non-parametric model Chao1 was chosen, which was the model that provided the best results for the datasets. Heatmaps were generated using the heatmaps from the ggplot2 R graphics library (Kolde, 2015).

2.4 Targeted metagenomics for the characterisation of adenovirus

To detect and characterise all mastadenoviruses and other potential adenoviruses present in raw sewage, general primers for AdV *hexon* were designed. To do so, the *hexon* region from 149 AdV genomes, recognised by the adenovirus taxonomy group and retrieved from GenBank, were analysed. The *hexon* region was selected based on its versatility as a very conserved/variable region (Hernroth et al., 2002) Due to the specific requirements of the Roche 454 Junior GS protocol, the designed primers were flanked with an adaptor and key sequences to identify the samples. The primers and conditions for *Adenovirus* PCR are presented in Supplementary material 3. PCR product was purified using the Zymo clean and concentrator

(cat no. D4013, Zymo Research). Purified amplicons were then pyrosequenced in a 454 GS Junior System (Life Science-Roche). Obtained raw reads in SFF were transformed to FASTQ using sff_extract from Roche. Adaptors were removed using Cutadapt (Martin, 2011) ; the complexity and quality of the reads were assessed by PrintSeq and FastQC (Schmieder and Edwards, 2011), and the reads were then trimmed using the FASTX-Toolkit software, version 0.0.14 (Hannon Lab). To define the non-redundant Operational Taxonomic Units (OTUs), CD-Hit was used and tested at different distance levels from which 0.02 was chosen. A local database was built that contained the *hexon* region of 153 adenovirus genomes available from GenBank (2016) and representing different species within the 5 *Adenoviridae* genera: *Aviadenovirus* (9), *Atadenovirus* (12), *Mastadenovirus* (122), *Siadenovirus* (4) and *Ichtadenovirus* (1). OTUs that matched the 0.02 criteria were blasted against the adenovirus local database using BLASTN (Altschul et al., 1997, 1990). A phylogenetic tree using Raxml with 1000 bootstrap replicates was computed using Geneious (Geneious 9,Kearse et al., 2012).

2.5. Virus amplification by experimental infection

In collaboration with Dr. Robert H. Purcell (Hepatitis Viruses Section, Laboratory of Infectious Diseases, NIAID, NIH, USA), experimental infections of two rhesus macaques (*Macaca mulatta*) that were previously immunised for hepatitis A virus (HAV) were carried out as part of a wider study at Bioqual in Rockville, MD, in compliance with the guidelines of the Institutional Animal Care and Use Committees of Bioqual and NIAID. The rhesus macaques were inoculated intravenously with 27 mL of 0.45-µm filtered raw sewage from Barcelona mixed with 3 mL of 10X PBS. Blood from both rhesus macaques was extracted on a weekly basis over two months to study the potential replication of human viruses present in raw sewage. A blood sample, used as a negative control, was extracted from each animal one week before the inoculation of raw sewage.

Sera samples were processed according to section 2.1.2 and libraries prepared according to section 2.2. In total, the following 4 different library preparations were sequenced: a pooled library prior to the raw sewage inoculation from the two rhesus monkeys (PW1), two different libraries from each of the animals one week after the inoculation (RW1 and RW2), and a pooled library from both rhesus monkeys 4 weeks after the inoculation (RW4).

3. Results and discussion

3.1. Mi-Seq run outputs in 10-L sewage samples from 3 different seasons

The Mi-Seq results obtained for the sequenced samples are summarised in Supplementary material 1. The viromes of the urban sewage collected during three different seasons—winter, spring and summer—were analysed using 10-L of raw sewage, and 37 different viral families were identified. The numbers of the different viral species assigned to a given viral family are graphically presented in Figure 2.

Bacteriophage families *Siphoviridae, Myoviridae, Podoviridae* and *Microviridae* show a higher diversity degree in urban sewage, which agrees with Clokie et al., 2011. The ssDNA parvoviruses, closely followed by the picornaviruses, constitute a diverse viral family whose members infect animals and humans. Viral plant *Virgaviridae* species are also abundantly represented in the samples. Important human viral pathogens that are taxonomically assigned to *Astroviridae, Caliciviridae*, *Hepeviridae* and *Polyomaviridae* were also detected. Furthermore, reads related to viruses belonging to the *Circoviridae* and *Picobirnaviridae* families were sequenced. A summary of the number of reads and contigs associated with these viral families can be found in Table 1. A complete list of detected viral sequences is provided as Supplementary material 2.

A wide diversity and abundance of human and animal astroviruses were detected in the winter sample. The majority of the reads from this sample belonged to the MAstV-1 genogroup, whereas MAstV-6, -8 and -9 were less frequent. Similarly, more sequences that were taxonomically assigned to the *Caliciviridae* viral family and that were specifically assigned to different norovirus GI and GII species and human sapoviruses were detected in winter. The seasonality of the astroviruses and caliciviruses during low-temperature seasons has been well-documented (Bosch et al., 2014; Haramoto et al., 2006). Within the *Picornaviridae* family, several human and animal picornaviruses were sequenced, including the recently described human salivirus/klassevirus, several Aichi viruses, and the recently described genus *Cosavirus*. The Aichi virus read counts were higher during summer compared to the other tested seasons. Human enteroviruses from species A, B, C and D had similar numbers, regardless of the analysed season. Important viral pathogens that cause hepatitis that is transmitted through the consumption of water/food contaminated with faecal material, such as HAV and hepatitis E virus (HEV), were only detected in low numbers in the winter sample. This can be related to the low prevalence of these infections in the studied area.

The viral faecal markers present in urban raw sewage, such as human adenoviruses, were not detected by metagenomics when the 10 L SMF protocol was applied. This contrasts with the detection of human adenoviruses by conventional qPCR in samples during winter, spring and summer, with concentration of $3.18x10^4$ genomic copies (GC)/L, $5.32x10^5$ GC/L and $1.23x10^5$ GC/L, respectively.

3.1.2. Targeted metagenomics for adenovirus characterisation

To address the lack of HAdV and to study the diversity of the genus *Mastadenovirus* in raw sewage, a target enrichment assay using broadly degenerate primers for the *hexon* region was conducted. Previously, concentrated SMF from spring was used because it contained higher genome copies of HAdV. A total of 55,903 raw reads were generated by pyrosequencing. All raw reads passed the cleaning cut-offs and were used for subsequent analyses. A sequence similarity of 98% was chosen as the cut-off for the homology searches, which resulted in a total of 3,677 different OTUs, accounting for 52,370 sequences from the sample (93.7%). The obtained OTUs were blasted against the custom-built adenovirus database, falling into 52 phylogenetically different AdV taxons. The detected AdVs from raw sewage are shown in Figure 3, and a complete list detailing the abundance of the detected AdVs is available in Supplementary material 3. Most of the sequences were assigned to murine adenovirus 2 (60%) and HAdV from species F, including HAdV-41 (29%) and HAdV-40 (0.7%). In total, 20 human adenovirus from species A, B, C, D, F and G were detected. The degenerate primers facilitated the detection of a wide range of AdVs with a high variability of hosts. However, given that some of the detected sequences were from AdV exotic animals and that they clustered with other well-known HAdV species, the used AdV database might not reflect the true diversity within the *Adenoviridae* family, and other excreted human/non-human adenoviruses may still need to be discovered. This is exemplified by several of the detected simian adenoviruses (SAdVs), which are closely related to HAdV40 and 41 (see Figure 3). Therefore, the detected SAdVs could be variants of the closely related HAdV40 and 41. It should also be considered that in the analysed short region, few changes are important, and errors may be introduced during the PCR amplification and sequencing process; a study by Niklas et al (2013) showed that 454 GS Junior has an overall error rate of 0.18%, and it is also known that the distribution of errors in the sequences is not homogenous.

3.2. Comparative evaluation of the ultracentrifugation and small-volume SMF protocols for viral concentration in sewage

To increase the detection sensitivity, two protocols were comparatively evaluated to concentrate viruses from sewage usingmetagenomics : a modified concentration protocol based on SMF with a lower sample volume (500 mL) and a protocol based on ultracentrifugation (42 mL). Compared with the results obtained using 10 L of urban sewage, the modified flocculation protocol allowed the detection of viral members and families previously not detected, despite the smaller sample volume tested. For example, *Adenoviridae, Polyomaviridae* and *Papillomaviridae* were identified when using smaller volumes.

Ultracentrifugation is an efficient technique to concentrate viruses, yielding good recoveries. However, difficulties in simultaneously concentrating viral particles from several samples and the requirement for an ultracentrifuge device hampers its applicability. A recent comparative study published in collaboration with Hjelmsø et al., (2017) showed that the analysis of 10 L of SMF, as described in section 2.1.1, in combination with QIAgen extraction columns (Iker et al., 2013) had inhibition problems, as evidenced by HAdV qPCR quantifications and the lower detection of HAdV by metagenomics. This observed inhibition might have affected the subsequent detection of several viral species by NGS. A simplified version of SMF using a smaller volume avoiding the ultracentrifugation step was compared against the reference ultracentrifugation protocol developed by Pina et al., (1998) to improve and minimise the observed limitations of the reference protocol. Both protocols assayed the same two collected sewage samples, testing the same volume of 7 raw sewage millilitre equivalents per library.

This comparative study allowed the detection of a wide variety of RNA and DNA pathogens, with a light increase in the number of sequences (principally bacteriophages) when using ultracentrifugation. The Mi-seq results are summarised in Supplementary material 4. A higher estimated viral richness is observed when using ultracentrifugation compared to SMF. This higher estimated viral richness values are explained by the high number of bacteriophages sequences detected using ultracentrifugation (see Figure 4). The four viral concentrates were analysed for HAdV by qPCR showing $8.14x10^5$ GC/L, $1.23x10^5$ GC/ L, $2.19x10^5$GC/L, and $1.48x10^5$ GC/L for HAdV in SMF1, SMF2, Ultra1, and Ultra2, respectively. In total, 41 different viral families were detected among all samples. A complete list of the detected viral families is highlighted in Figure 4. The modified SMF protocol, with the reduction in the sample volume, allowed the detection of 36 different viral families compared to the 38 different viral families detected by ultracentrifugation. Few viral human species showing a low number of contigs were detected only by the ultracentrifugation protocol, such as the *Anelloviridae*, *Alloherpesviridae, Geminiviridae, Hepeviridae*, *Totiviridae, Geminiviridae,* and *Polyomaviridae* families. Other viral families, such as the *Luteoviridae, Nanoviridae,* and *Baculoviridae* families, were only detected

using SMF. For most of the important viral families, including human pathogenic viruses such as *Adenoviridae, Caliciviridae, Parvoviridae, Circoviridae, Astroviridae, and Picornaviridae,* a high diversity of viral species were detected with similar results by both the SMF-500 mL and ultrafiltration protocols, demonstrating the suitability of these concentration methods for the detection of pathogens such as caliciviruses, the main viruses responsible for gastroenteritis outbreaks (Ahmed et al., 2014). Hence, the availability of an effective concentration method to detect pathogenic viruses is crucial if NGS metagenomic data will be used for surveillance purposes. The efficacy of SMF to concentrate ssRNA+ viral particles agrees with previously published results by Hjelmsø et al.(2017). The low-volume SMF protocol allowed for the detection of a previously undetected family, *Adenoviridae.* Human adenoviruses detected by untargeted metagenomics were taxonomically assigned to the human adenovirus F species (HAdV-40 and HAdV-41). The results obtained in urban sewage using the specific adenovirus targeted metagenomics assay showed a wide diversity of adenoviruses, including up to 20 Human adenoviruses. Murine adenovirus 2 was found to be the most abundant *Adenoviridae* representative in this specific sample analysed, and HAdV-40, HAdV-41 and low numbers of other adenoviral species were also detected. The specific characteristics of the sample and a possible biased preference for the murine adenoviruses of the highly degenerated adenovirus *hexon* primers used in the targeted assay may partially explain the high number of sequences assigned to this viral species.

Larger analysed sample volumes (10 L vs 500 mL) may increase the chances of detecting rare viral families in sewage. However, larger volumes also have a higher proportion of inhibitors (Schrader et al., 2012). Inhibitors might have affected the PCR amplification step, considering that 40 cycles were needed for the 10-L SMF to prepare the libraries, while only 25 amplification cycles were needed when 500 mL of SMF was used. Viral metagenomics is limited by the low levels of viral DNA/RNA present in the samples, in most cases requiring a PCR-based random amplification step after the RT and Sequenase reactions to obtain sufficient DNA for library preparation. Interestingly, the viral richness was quite similar despite the different PCR amplification cycles applied. The PCR amplification step might introduce bias by amplifying the most abundant genomes such that less abundant genomes might not be sequenced or may be underrepresented (Karlsson et al., 2013). This might be the case for HAdV, as data obtained in previous assays showed that PCR random amplification methods decreased the estimated viral richness of the dsDNA genomes more significantly compared to other viral genomes (data not shown). Overall, the data indicate that a concentration of 500 mL of urban raw sewage constitutes a representative sample volume to study the virome of raw sewage.

One of the main objectives of this research was to shed light on the viral families present in raw sewage, which we define as the sewage virome. This list should be periodically reviewed using the developed protocols for environmental surveillance and to identify the introduction of pathogens and novel or emerging viral strains in the population and environment. A complete list of the different viral species detected in raw sewage in this study is provided in Supplementary material 2.

In total, more than 11 different viral families considered, or putatively considered, as pathogenic have been detected in raw sewage from Barcelona. Human astroviruses (HAstV) are suspected to be involved in 0.5 to 15% of all acute diarrhoea outbreaks in children (Bosch et al., 2014). In the present study, a high diversity of sequences, mainly assigned to the MastV-1 genotype, was detected in all tested samples, but several recombinant genotypes, such as MAstV-6, -8 and -9, were observed in lower abundance. More precisely, the application of NGS techniques has facilitated the detection of these later mentioned animal recombinant astroviruses (Finkbeiner et al., 2009; Kapoor et al., 2009), which are related to neurological disorders in immunocompromised patients (Brown et al., 2015).

NoV from both genogroups GI and GII were detected in all sewage samples, reflecting a wide diversity within this variable viral family. The included sequences were taxonomically assigned to NoV GII.4 and NoV GII.17, which are the more frequently reported gastroenteritis genotypes (Chan et al., 2015; Vega et al., 2011). Within the same family, human sapoviruses (HSaV) that belong to GI, GII, GIV, and GV were also found; they were previously reported as gastroenteritis agents (Oka et al., 2015).

Picornaviridae is a family that comprises more than 30 different genera of ssRNA+ viruses including important human pathogens, such as the hepatitis A virus and poliovirus. Several species of the genera *Kobuvirus*, *Enterovirus (EV)*, *Cosavirus, Salivirus,* and *Cardiovirus* were detected in sewage. Aichi virus (AiV) has been recovered during all seasons and in all tested sewage samples, which agrees with the available data (Lodder et al., 2013). Recent studies have suggested that AiV may co-infect with other enteric viruses, causing gastroenteritis (Ambert-Balay et al., 2008; Räsänen et al., 2010). EV is one of the most important genera within the *Picornaviridae* family; EV contains 12 different species that infect humans, including EV species A to D and *Rhinovirus* species A to C (Plyusnin et al., 2011). Different EV from species A, B, and C and animal enteroviruses from species G and J were also noted. Most of the identified human enteroviruses belong to species A and B, but important enteroviruses from species C, such as Enterovirus-A71, were detected. An increase in enterovirus outbreaks has recently been reported to be caused by emerging recombinant EV strains (Holm-Hansen et al., 2016; Zhang et al., 2010).

Other sequences related to the *Salivirus* and *Cosavirus* genera, whose causal role in gastroenteritis is suspected, have been detected (Li et al., 2009; Tseng et al., 2007).

Several sequences resembling animal parvoviruses that infect dogs, rats, cattle and swine and several densoviruses that infect invertebrates have been identified. Human bocavirus (HBoV) species HBoV1, 2, 3 and, 4 and human bufaviruses have been observed, yet the implications of these parvoviruses in human disease are controversial (Nawaz et al., 2012; Phan et al., 2012), and further studies need to be conducted to better characterise their pathogenic roles or consider them as part of the human gut viral community.

Sequences that are taxonomically assigned to the *Circoviridae* family have been detected in all sewage samples. Because circoviruses are prevalent in several human fluids, their detection in raw sewage seems reasonable. Their relationship with disease remains unclear, but cycloviruses are involved in acute nervous system infections (Tan et al., 2013).

Orthohepevirus, within the *Hepeviridae* family*,* is the genus of the species Orthoherpesvirus A, which includes the viruses causing hepatitis in humans. Genotypes 1 and 2 have been reported to infect only humans, while genotypes 3 and 4 are zoonotic (Legrand-Abravanel et al., 2009). The finding in one sample of HEV genotype 3, frequently detected in swine, demonstrates a low prevalence compared to other faecally transmitted viruses causing gastroenteritis (Rutjes et al., 2014).

Surprisingly, no members of the *Reoviridae* family were detected. Important pathogenic viruses within this family include the human rotaviruses, which are important gastroenteritis agents in children that caused approximately 453,000 deaths in 2008 (Tate et al., 2012). Although rotaviruses are detected in similar concentrations compared to other enteric viruses in sewage (Prado et al., 2011), their prevalence is lower compared to HAdV and influenced by seasonality patterns (El-Senousy et al., 2015; Zhou et al., 2016). Other metagenomic studies failed to detect rotaviruses, although they included sewage samples from endemic rotavirus areas (Cantalupo et al., 2011; Ng et al., 2012). *Picobirnaviridae* viruses from the family*,* which also have dsRNA segmented genomes, have been detected in all tested raw sewage samples. Human *Picobirnaviruses* are prevalent by conventional PCR in 100% of sewage samples and have been detected at high concentrations (Symonds et al., 2009). Again, a higher relative abundance of this viral family compared to rotaviruses should be expected.

In the present study, dsDNA viral families, such as *Polyomaviridae*, *Adenoviridae* and *Papillomaviridae*, have been detected. Polyomaviruses and adenoviruses are excreted by symptomatic and asymptomatic carriers, independent of the seasonality or geographical area. Therefore, they are present in nearly 100% of untreated sewage, which makes them suitable as human viral faecal indicators (Bofill-Mas et al., 2013). Human papillomaviruses (HPV) have

recently been reported in raw sewage (La Rosa et al., 2013). The transmission of papillomaviruses through the consumption of faecal contaminated water or food remains unproven, and further studies on the significance of their molecular detection are needed. Families with insect viruses, such as *Dicistroviridae*, *Iridoviridae,* and *Nodaviridae*, have also been detected; insects can be expected to enter through the sewage system of a city.

A high abundance and diversity of plant viruses was found in our samples. Viruses from the *Virgaviridae, Closteroviridae, Partitiviridae, Alphaflexiviridae, Betaflexiviridae, Tombusviridae, Bromoviridae, Secoviridae, Potyviridae,* and *Tymoviridae* families seem to be abundant and important components of the sewage virome. The members of the *Virgaviridae* family were especially diverse; they were the second most diverse family to be detected regardless of the concentration method or volume. Plant viruses are highly abundant in human faeces (Zhang et al., 2006). For example, PMMV has been recently connected to specific immune responses, fever, and abdominal pain in humans by Colson et al. (Colson et al., 2010). The infectivity of human excreted plant viruses has already been demonstrated (Tomlinson et al., 1982; Zhang et al., 2006). As a result, their presence in STP effluents may represent an economic threat for farmers if reclaimed water without suitable quality control is used for crop irrigation.

Bacteriophages composed the major fraction from the sewage virome, with sequences identified from the *Microviridae*, *Podoviridae*, *Myoviridae*, *Leviviridae* and *Siphoviridae* families. *Microviridae* is the family with the highest level of diversity. The detected phage viral families in the present study agree with other untargeted metagenomic analyses (Tamaki et al., 2012). It is likely that the number of bacteriophage sequences has been underestimated due to the taxonomical assignment of prophages as bacterial DNA.

The application of NGS techniques to environmental and clinical samples facilitates the simultaneous analysis of millions of sequences. Of note, a significant fraction of sequences remains unassigned to known taxonomic units after bioinformatics analyses. In the present study, samples were virion-enriched by the applied concentration methods, and the viral concentrate was filtered to remove bacteria, while DNase was used to remove free DNA. Nevertheless, the percentage of sequences assigned to a known virus taxon was extremely low, but the data agreed with previous publications.

The evaluated sewage virome is only an initial attempt to address complex water matrices. The lack of a universal viral marker-compared to bacterial 16S and the need to sequence all available RNA/DNA present in samples requires concentration methods for viral particles while removing other DNA sources to increase the sensitivity of viral metagenomics. It is expected that the

development and availability of improved sequencing technologies, such as single-molecule nanopore sequencers, in the forthcoming years will provide a more accurate and detailed composition description of the viral mixtures from different types of samples, including those of the sewage virome.

The annotation of the urban sewage virome using NGS methods describes the catalogue of the viral species circulating across a given population, which increasingly plays an important role in public health surveillance. Viruses are more resistant than bacteria to specific treatments applied in STPs. Therefore, they can be present in reclaimed water produced for crop irrigation, surpassing FIB microbiological quality parameters. A previous study by Rosario et al., (2009) demonstrated that reclaimed water contains 1000-fold more virus-like particles than potable water. Although no pathogenic viruses were detected in that study, pathogenic infectious viruses have been detected in reclaimed water in other studies (Rodriguez-Manzano et al., 2012).In the later study, infective Human adenovirus, were detected in the reclaimed water produced, highlighting that improved treatments designed for viral removal should be implemented to produce safer reclaimed water. For the protection of public health, true viral indicator or indicators should be pursued with further studies considering that the methods for concentration and quantification would need to be harmonised. Different viruses have been proposed as viral indicators, including the highly abundant viruses detected in this study, Aichivirus (a picornavirus from kobuvirus genus (Kitajima and Gerba, 2015; Lodder et al., 2013), picobirnavirus (Symonds et al., 2009), and a highly abundant plant virus PMMV (Kuroda et al., 2015; K. Rosario et al., 2009), in addition to coliphages that have also limitations but are considered useful indicators for the availability of culture standard methods. The HAdV have been proposed and used as a viral fecal indicator from human origin due to their high abundance in all seasons, stability in the environment and resistance to the STP treatments commonly applied, and their importance in public health for their pathogenicity. HAdV monitoring in reclaimed water produced could be useful to increase water safety. Important viral pathogens, specially HNoV, should be monitored by molecular methods, however, the lack of an infectivity model hampers their applicability to monitor water quality as there is not model to extrapolate genome copies to infectivity if most of the complex water treatments used in reclaimed water are considered. The results derived from this study point out that a high diversity of known and unknown viral families are present in raw sewage. The list of viruses that should be explored to ensure water reuse safety should be updated and revised.

3.3. The contribution of urine to the viral composition of sewage

Detected viral sequences from the human urine samples analysed are summarised in Supplementary material 5. The urine viral concentrate contained the following DNA viral families that infect humans: *Papillomaviridae, Polyomaviridae,* and sequences distantly related to circular ssDNA families *Circoviridae,* and *Anelloviridae*. These results highlight that urine contributes to the highly diverse viral composition of urban sewage by introducing primarily DNA viruses. Human polyomaviruses, namely, the JC polyomaviruses (JCPyVs) that are known to be excreted through urine mainly, were the most abundant; the BK polyomaviruses (BKPyV) showed a lower number of sequences, where 0.76% of the total reads were associated with this family. This excretion route for polyomaviruses has already been documented in the literature (Egli et al., 2009; Shinohara et al., 1993). For this reason, the group has been widely used as a specific indicator of human excreta in water (Harwood et al., 2009). In recent years, new polyomaviruses have been described, including up to 13 human polyomaviruses (Mishra et al., 2014). MCPyV is not excreted through urine (Loyo et al., 2010). Instead, it is frequently detected in skin samples in conjunction with human Polyomaviruses 6, 7, and 9 (Foulongne et al., 2012). The lack of detection of the new polyomavirus from urine samples suggests that the excretion patterns of these polyomaviruses might occur through faeces and skin desquamation. Reads for HPV (0.03% of total reads), matching HPV129 and HPV170, were identified; the HPVs probably originated from epithelial desquamation during urination. The detection of HPVs has been reported in faeces (Di Bonito et al., 2015), raw sewage (La Rosa et al., 2013), and urine (Santiago-Rodriguez et al., 2015). In a prior study, several β-HPV (HPV49, HPV92, and HPV96) and γHPV (HPV121 and HPV178) samples were detected. HPV species detected in this study have not been reported in any of the urine metagenomic studies available to date (Santiago-Rodriguez et al., 2015; Smelov et al., 2016, 2014). Although skin desquamation and excretion through faeces might be the main modes through which human papillomaviruses arrive in sewage, the excretion of specific papillomaviruses such as the skin-specific γHPV, which might have tropism for the urinary tract, is a notable finding. Because none of the volunteers participating in this study had been diagnosed with HPV infections or genital warts, HPV may be part of the virome of the urinary tract without causing any known disease. More urine-focused studies, including investigations that apply specific PCR target enrichment to sequencing, would improve our knowledge of the diversity of HPVs in urine. Sequences that are distantly related at the protein level to *Circoviridae* and *Anelloviridae* were also observed. ssDNA viruses seem to be ubiquitously present in blood (Vasilyev et al., 2009). Therefore, the detection of these specific

viral families in urine seems very plausible. With the advent of NGS techniques, there has been a significant increase in viruses classified under these two ssDNA viral families (Kim et al., 2011) and other ssDNA circular viral particles that remain unclassified (Kim et al., 2012).

3.4. Identified infective human viruses present in raw sewage amplified by experimental infection

One week after inoculation, the first rhesus monkey presented reads matching JCPyV and the hepatitis E virus. This observation is consistent with the active replication of these two human viruses identified in raw sewage using animal models. The HEV strain found in the rhesus blood sample was annotated as genotype 3. The inoculation of environmental HEV strains into rhesus monkeys is an effective method to replicate the virus (Pina et al., 1998a). Sequences classified within this genotype are frequently reported in the geographical area of the study, i.e., Europe (Clemente-Casares et al., 2009), and this genotype is one of the most commonly detected HEV genotypes in Europe and North America (Clemente-Casares et al., 2003). The second rhesus monkey did not present JCPyV or HEV sequences in its serum on the studied dates (one week and one month after inoculation). The pooled sample from both rhesus monkeys at 4 weeks post-inoculation did not contain any sequences related to the Hepatitis E virus or JCPyV, supporting the model of an acute asymptomatic infection. The pooled serum samples collected one week before the inoculation showed the presence of several viral plants from the *Virgaviridae* family and several phages from the *Microviridae* and *Inoviridae* families. Large fractions of genomic plant DNA have been detected in blood (Spisák et al., 2013), suggesting the possibility that viral DNA/RNA could also be circulating through blood and thus be detected by metagenomics. A total of 1,462 reads (0.08%) in sera samples after inoculation were taxonomically assigned to the *Anelloviridae* family and more specifically to human Torque teno viruses (TTVs) 26 and 27. These two viral species were detected in all rhesus serum samples, supporting the wide distribution and prevalence of these viruses among mammals (de Villiers and Hausen, 2009). The presence of these viruses in blood has also been reported in humans without any associated disease (Biagini et al., 2013).

Conclusions

Raw sewage harbours a vast number of different viral families that may contaminate the environment since typically, viruses are not completely removed in STPs. The methodologies developed based on ultracentrifugation and, if an ultracentrifuge is unavailable, the SMF protocol for 500-mL samples are useful and produce robust results for the characterisation of

the virome of urban sewage by detecting both DNA and RNA viruses. Virome information for urban sewage may constitute an important database for known, novel and emerging viral strains that are excreted in the population at a specific time. Among the human viral families, important human pathogens have been detected by NGS, including members of the *Parvoviridae, Caliciviridae, Hepeviridae, Adenoviridae, Polyomaviridae, Papillomaviridae, Picornaviridae* and *Astroviridae* families. The implementation and application of a low-volume SMF protocol minimises the inhibition problems detected when sampling larger volumes, while offering a representative volume that yields comparable results to those of the ultracentrifugation method. However, the sensitivity for analysing specific viral groups and the reduction of representation biases for relatively less abundant viral species is increased when targeted metagenomics assays that are designed to amplify specific viral species are used.

The use of experimentally infected rhesus macaques for the amplification of viruses excreted in sewage enabled the detection of infective HEV and JCPyV from urban sewage. This approach also provided important information regarding the presence of plant viruses in the sera of the macaques and the presence of small unclassified cDNA viruses, which will merit further studies.

The contribution of urine to sewage seems limited to DNA viral families, mainly JCPyVs, which appear to be highly excreted, and BKPyV, at lower quantities.

The use of NGS techniques for sewage analysis can pinpoint major pathogens that circulate in the population and environment, making NGS techniques useful tools for epidemiologic studies and public health surveillance.

Acknowledgements

We would like to acknowledge R.H. Purcell from the NIH (Bethesda, USA) for the Rhesus macaque experiments and providing the Rhesus sera samples. This study was partially funded by the Programa RecerCaixa 2012 (ACUP-00300) the projects AGL2011-30461-C02-01/ALI from the Spanish Ministry of Science and Innovation and the project AGL2014-55081-R from the Ministry of Economy. This study was partially funded by grants from the Catalan Government to the Consolidated Research Group VirBaP (2014SRG914) and JPI Water project METAWATER (4193-00001B), with the collaboration of the Institut de Recerca de l'Aigua (IdRA) and the Spanish Adenovirus Network (AdenoNet, BIO2015-68990-REDT) from the Ministerio de Economía y Competitividad of Spain (www.mineco.gob.es). During the development of this study, Xavier Fernandez-Cassi was a fellow of the Catalan Government "AGAUR" (FI-DGR); Natalia Timoneda and Sandra Martínez-Puchol were fellows of the Spanish Ministry of Science.

Bibliography

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. doi:10.1016/S1473-3099(14)70767-4

Allen, H.K., Bunge, J., Foster, J.A., Bayles, D.O., Stanton, T.B., 2013. Estimation of viral richness from shotgun metagenomes using a frequency count approach. Microbiome 1, 5. doi:10.1186/2049-2618-1-5

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–10. doi:10.1016/S0022-2836(05)80360-2

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. doi:10.1093/nar/25.17.3389

Ambert-Balay, K., Lorrot, M., Bon, F., Giraudon, H., Kaplon, J., Wolfer, M., Lebon, P., Gendrel, D., Pothier, P., 2008. Prevalence and genetic diversity of Aichi virus strains in stool samples from community and hospitalized patients. J. Clin. Microbiol. 46, 1252–1258. doi:10.1128/JCM.02140-07

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2015. GenBank. Nucleic Acids Res. 43, D30–D35. doi:10.1093/nar/gku1216

Biagini, P., Bédarida, S., Touinssi, M., Galicher, V., de Micco, P., 2013. Human gyrovirus in healthy blood donors, France. Emerg. Infect. Dis. 19, 1014–5. doi:10.3201/eid1906.130228

Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Appl. Environ. Microbiol. 72, 7894–6. doi:10.1128/AEM.00965-06

Bofill-Mas, S., Rusiñol, M., Fernandez-Cassi, X., Carratalà, A., Hundesa, A., Girones, R., 2013. Quantification of human and animal viruses to differentiate the origin of the fecal contamination present in environmental samples. Biomed Res. Int. 2013, 192089. doi:10.1155/2013/192089

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170

630 Bosch, A., Pintó, R.M., Guix, S., 2014. Human astroviruses. Clin. Microbiol. Rev. 27, 1048–1074.
631 doi:10.1128/CMR.00013-14

632 Brister, J.R., Ako-Adjei, D., Bao, Y., Blinkova, O., 2015. NCBI viral genomes resource. Nucleic
633 Acids Res. 43, D571-7. doi:10.1093/nar/gku1207

634 Brown, J.R., Morfopoulou, S., Hubb, J., Emmett, W.A., Ip, W., Shah, D., Brooks, T., Paine, S.M.L.,
635 Anderson, G., Virasami, A., Tong, C.Y.W., Clark, D.A., Plagnol, V., Jacques, T.S., Qasim, W.,
636 Hubank, M., Breuer, J., 2015. Astrovirus VA1/HMO-C: an increasingly recognized
637 neurotropic pathogen in immunocompromised patients. Clin. Infect. Dis. 60, 881–8.
638 doi:10.1093/cid/ciu940

639 Calgua, B., Mengewein, A., Grunert, A., Bofill-Mas, S., Clemente-Casares, P., Hundesa, A., Wyn-
640 Jones, A.P., López-Pila, J.M., Girones, R., 2008. Development and application of a one-step
641 low cost procedure to concentrate viruses from seawater samples. J. Virol. Methods 153,
642 79–83. doi:10.1016/j.jviromet.2008.08.003

643 Cantalupo, P.G., Calgua, B., Zhao, G., 2011. Raw Sewage Harbors Diverse Viral Populations 2, 1–
644 11. doi:10.1128/mBio.00180-11.Editor

645 Chan, M.C.W., Lee, N., Hung, T.-N., Kwok, K., Cheung, K., Tin, E.K.Y., Lai, R.W.M., Nelson, E.A.S.,
646 Leung, T.F., Chan, P.K.S., 2015. Rapid emergence and predominance of a broadly
647 recognizing and fast-evolving norovirus GII.17 variant in late 2014. Nat. Commun. 6,
648 10061. doi:10.1038/ncomms10061

649 Clemente-Casares, P., Pina, S., Buti, M., Jardi, R., MartIn, M., Bofill-Mas, S., Girones, R., 2003.
650 Hepatitis E virus epidemiology in industrialized countries. Emerg. Infect. Dis. 9, 448–54.
651 doi:10.3201/eid0904.020351

652 Clemente-Casares, P., Rodriguez-Manzano, J., Girones, R., 2009. Hepatitis E virus genotype 3
653 and sporadically also genotype 1 circulate in the population of Catalonia, Spain. J. Water
654 Health 7, 664. doi:10.2166/wh.2009.120

655 Clokie, M.R., Millard, A.D., Letarov, A. V, Heaphy, S., 2011. Phages in nature. Bacteriophage 1,
656 31–45. doi:10.4161/bact.1.1.14942

657 Colson, P., Richet, H., Desnues, C., Balique, F., Moal, V., Grob, J.-J., Berbis, P., Lecoq, H., Harlé, J.-
658 R., Berland, Y., Raoult, D., 2010. Pepper Mild Mottle Virus, a Plant Virus Associated with
659 Specific Immune Responses, Fever, Abdominal Pains, and Pruritus in Humans. PLoS One 5,
660 e10041. doi:10.1371/journal.pone.0010041

661 de Villiers, E.-M., Hausen, H. zur (Eds.), 2009. TT Viruses, Current Topics in Microbiology and
662 Immunology. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-
663 70972-5

664 Di Bonito, P., Della Libera, S., Petricca, S., Iaconelli, M., Sanguinetti, M., Graffeo, R., Accardi, L.,
665 La Rosa, G., 2015. A large spectrum of alpha and beta papillomaviruses are detected in
666 human stool samples. J. Gen. Virol. 96, 607–613. doi:10.1099/vir.0.071787-0

667 Egli, A., Infanti, L., Dumoulin, A., Buser, A., Samaridis, J., Stebler, C., Gosert, R., Hirsch, H.H.,
668 2009. Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood
669 donors. J. Infect. Dis. 199, 837–46.

670 El-Senousy, W.M., Ragab, A.M.E.-S., Handak, E.M.A.E.H., 2015. Prevalence of Rotaviruses
671 Groups A and C in Egyptian Children and Aquatic Environment. Food Environ. Virol. 7, 132–
672 141. doi:10.1007/s12560-015-9184-6

673 Finkbeiner, S.R., Li, Y., Ruone, S., Conrardy, C., Gregoricus, N., Toney, D., Virgin, H.W., Anderson,
674 L.J., Vinjé, J., Wang, D., Tong, S., 2009. Identification of a novel astrovirus (astrovirus VA1)
675 associated with an outbreak of acute gastroenteritis. J. Virol. 83, 10836–9.
676 doi:10.1128/JVI.00998-09

677 Foulongne, V., Sauvage, V., Hebert, C., Dereure, O., Cheval, J., Gouilh, M.A., Pariente, K.,
678 Segondy, M., Burguière, A., Manuguerra, J.-C., Caro, V., Eloit, M., 2012. Human Skin
679 Microbiota: High Diversity of DNA Viruses Identified on the Human Skin by High
680 Throughput Sequencing. PLoS One 7, e38499. doi:10.1371/journal.pone.0038499

681 Gerba, C.P., Goyal, S.M., LaBelle, R.L., Cech, I., Bodgan, G.F., 1979. Failure of indicator bacteria
682 to reflect the occurrence of enteroviruses in marine waters. Am. J. Public Health 69, 1116–
683 9.

684 Haramoto, E., Katayama, H., Oguma, K., Yamashita, H., Tajima, A., Nakajima, H., Ohgaki, S.,
685 2006. Seasonal profiles of human noroviruses and indicator bacteria in a wastewater
686 treatment plant in Tokyo, Japan. Water Sci. Technol. 54, 301–8.

687 Harwood, V.J., Brownell, M., Wang, S., Lepo, J., Ellender, R.D., Ajidahun, A., Hellein, K.N.,
688 Kennedy, E., Ye, X., Flood, C., 2009. Validation and field testing of library-independent
689 microbial source tracking methods in the Gulf of Mexico. Water Res. 43, 4812–4819.
690 doi:10.1016/j.watres.2009.06.029

691 Hernroth, B.E., Conden-Hansson, A.C., Rehnstam-Holm, A.S., Girones, R., Allard, A.K., 2002.

692 Environmental factors influencing human viral pathogens and their potential indicator
693 organisms in the blue mussel, Mytilus edulis: The first Scandinavian report. Appl. Environ.
694 Microbiol. 68, 4523–4533. doi:10.1128/AEM.68.9.4523-4533.2002

695 Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M.,
696 Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz,
697 A.C., 2017. Evaluation of Methods for the Concentration and Extraction of Viruses from
698 Sewage in the Context of Metagenomic Sequencing. PLoS One 12, e0170199.
699 doi:10.1371/journal.pone.0170199

700 Holm-Hansen, C.C., Midgley, S.E., Fischer, T.K., 2016. Global emergence of enterovirus D68: A
701 systematic review. Lancet Infect. Dis. doi:10.1016/S1473-3099(15)00543-5

702 Iker, B.C., Bright, K.R., Pepper, I.L., Gerba, C.P., Kitajima, M., 2013. Evaluation of commercial kits
703 for the extraction and purification of viral nucleic acids from environmental and fecal
704 samples. J. Virol. Methods 191, 24–30. doi:10.1016/j.jviromet.2013.03.011

705 Kapoor, A., Li, L., Victoria, J., Oderinde, B., Mason, C., Pandey, P., Zaidi, S.Z., Delwart, E., 2009.
706 Multiple novel astrovirus species in human stool. J. Gen. Virol. 90, 2965–2972.
707 doi:10.1099/vir.0.014449-0

708 Karlsson, O.E., Belák, S., Granberg, F., 2013. The effect of preprocessing by sequence-
709 independent, single-primer amplification (SISPA) on metagenomic detection of viruses.
710 Biosecur. Bioterror. 11 Suppl 1, S227-34. doi:10.1089/bsp.2013.0008

711 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper,
712 A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012.
713 Geneious Basic: An integrated and extendable desktop software platform for the
714 organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
715 doi:10.1093/bioinformatics/bts199

716 Kim, H.K., Park, S.J., Nguyen, V.G., Song, D.S., Moon, H.J., Kang, B.K., Park, B.K., 2012.
717 Identification of a novel single-stranded, circular DNA virus from bovine stool. J. Gen. Virol.
718 93, 635–639. doi:10.1099/vir.0.037838-0

719 Kim, M.-S., Park, E.-J., Roh, S.W., Bae, J.-W., 2011. Diversity and abundance of single-stranded
720 DNA viruses in human feces. Appl. Environ. Microbiol. 77, 8062–70.
721 doi:10.1128/AEM.06331-11

722 Kitajima, M., Gerba, C.P., 2015. Aichi virus 1: environmental occurrence and behavior. Pathog.

(Basel, Switzerland) 4, 256–68. doi:10.3390/pathogens4020256

Kohl, C., Brinkmann, A., Dabrowski, P.W., Radonić, A., Nitsche, A., Kurth, A., 2015. Protocol for metagenomic virus detection in clinical specimens. Emerg. Infect. Dis. 21, 48–57. doi:10.3201/eid2101.140766

Kolde, R., 2015. pheatmap: Pretty Heatmaps.

Kuroda, K., Nakada, N., Hanamoto, S., Inaba, M., Katayama, H., Do, A.T., Nga, T.T.V., Oguma, K., Hayashi, T., Takizawa, S., 2015. Pepper mild mottle virus as an indicator and a tracer of fecal pollution in water environments: Comparative evaluation with wastewater-tracer pharmaceuticals in Hanoi, Vietnam. Sci. Total Environ. 506–507, 287–298. doi:10.1016/j.scitotenv.2014.11.021

La Rosa, G., Fratini, M., Accardi, L., D'Oro, G., Della Libera, S., Muscillo, M., Di Bonito, P., 2013. Mucosal and Cutaneous Human Papillomaviruses Detected in Raw Sewages. PLoS One 8, e52391. doi:10.1371/journal.pone.0052391

Legrand-Abravanel, F., Mansuy, J.-M., Dubois, M., Kamar, N., Peron, J.-M., Rostaing, L., Izopet, J., 2009. Hepatitis E virus genotype 3 diversity, France. Emerg. Infect. Dis. 15, 110–4. doi:10.3201/eid1501.080296

Li, L., Victoria, J., Kapoor, A., Blinkova, O., Wang, C., Babrzadeh, F., Mason, C.J., Pandey, P., Triki, H., Bahri, O., Oderinde, B.S., Baba, M.M., Bukbuk, D.N., Besser, J.M., Bartkus, J.M., Delwart, E.L., 2009. A novel picornavirus associated with gastroenteritis. J. Virol. 83, 12002–6. doi:10.1128/JVI.01241-09

Linsuwanon, P., Poovorawan, Y., Li, L., Deng, X., Vongpunsawad, S., Delwart, E., 2015. The fecal virome of children with hand, foot, and mouth disease that tested PCR negative for pathogenic enteroviruses. PLoS One 10. doi:10.1371/journal.pone.0135573

Lodder, W.J., Rutjes, S.A., Takumi, K., de Roda Husman, A.M., 2013. Aichi virus in sewage and surface water, the Netherlands. Emerg. Infect. Dis. 19, 1222–1230. doi:10.3201/eid1908.130312

Loyo, M., Guerrero-Preston, R., Brait, M., Hoque, M.O., Chuang, A., Kim, M.S., Sharma, R., Liégeois, N.J., Koch, W.M., Califano, J.A., Westra, W.H., Sidransky, D., 2010. Quantitative detection of Merkel cell virus in human tissues and possible mode of transmission. Int. J. Cancer 126, NA-NA. doi:10.1002/ijc.24737

Ly, M., Abeles, S.R., Boehm, T.K., Robles-Sikisaka, R., Naidu, M., Santiago-Rodriguez, T., Pride,

754 D.T., 2014. Altered oral viral ecology in association with periodontal disease. MBio 5,
755 e01133-14. doi:10.1128/mBio.01133-14

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

758 Mihalov-Kovács, E., Fehér, E., Martella, V., Bányai, K., Farkas, S.L., 2014. The fecal virome of
759 domesticated animals. Virusdisease 25, 150–7. doi:10.1007/s13337-014-0192-1

760 Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., Bushman, F.D., 2011.
761 The human gut virome: inter-individual variation and dynamic response to diet. Genome
762 Res. 21, 1616–25. doi:10.1101/gr.122705.111

763 Mishra, N., Pereira, M., Rhodes, R.H., An, P., Pipas, J.M., Jain, K., Kapoor, A., Briese, T., Faust,
764 P.L., Lipkin, W.I., 2014. Identification of a Novel Polyomavirus in a Pancreatic Transplant
765 Recipient With Retinal Blindness and Vasculitic Myopathy. J. Infect. Dis. 210, 1595–1599.
766 doi:10.1093/infdis/jiu250

767 Nawaz, S., Allen, D.J., Aladin, F., Gallimore, C., Iturriza-Gómara, M., 2012. Human bocaviruses
768 are not significantly associated with gastroenteritis: results of retesting archive DNA from
769 a case control study in the UK. PLoS One 7, e41346. doi:10.1371/journal.pone.0041346

770 Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S.,
771 Wommack, K.E., Delwart, E., 2012. High Variety of Known and New RNA and DNA Viruses
772 of Diverse Origins in Untreated Sewage. J. Virol. 86, 12161–12175. doi:10.1128/JVI.00869-
773 12

774 Niklas, N., Pröll, J., Danzer, M., Stabentheiner, S., Hofer, K., Gabriel, C., 2013. Routine
775 performance and errors of 454 HLA exon sequencing in diagnostics. BMC Bioinformatics
776 14, 176. doi:10.1186/1471-2105-14-176

777 Oka, T., Wang, Q., Katayama, K., Saif, L.J., 2015. Comprehensive review of human sapoviruses.
778 Clin. Microbiol. Rev. 28, 32–53. doi:10.1128/CMR.00011-14

779 Perlejewski, K., Popiel, M., Laskus, T., Nakamura, S., Motooka, D., Stokowy, T., Lipowski, D.,
780 Pollak, A., Lechowicz, U., Caraballo Cortés, K., Stępień, A., Radkowski, M., Bukowska-Ośko,
781 I., 2015. Next-generation sequencing (NGS) in the identification of encephalitis-causing
782 viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens.
783 J. Virol. Methods 226, 1–6. doi:10.1016/j.jviromet.2015.09.010

784 Phan, T.G., Vo, N.P., Bonkoungou, I.J.O., Kapoor, A., Barro, N., O'Ryan, M., Kapusinszky, B.,

785 Wang, C., Delwart, E., 2012. Acute Diarrhea in West African Children: Diverse Enteric
786 Viruses and a Novel Parvovirus Genus. J. Virol. 86, 11024–11030. doi:10.1128/JVI.01427-
787 12

788 Pina, S., Jofre, J., Emerson, S.U., Purcell, R.H., Girones, R., 1998a. Characterization of a strain of
789 infectious hepatitis E virus isolated from sewage in an area where hepatitis E is not
790 endemic. Appl. Environ. Microbiol. 64, 4485–8.

791 Pina, S., Puig, M., Lucena, F., Jofre, J., Girones, R., 1998b. Viral pollution in the environment and
792 in shellfish: human adenovirus detection by PCR as an index of human viruses. Appl.
793 Environ. Microbiol. 64, 3376–82.

794 Plyusnin, A., Beaty, B.J., Elliott, R.M., Goldbach, R., Kormelink, R., Lundkvist, A., Schmaljohn, C.S.,
795 Tesh, R.B., 2011. Virus Taxonomy: Ninth Report of the International Committee on
796 Taxonomy of Viruses, vol. 9, in: Edited by King, AMQ, Adams, J. & Lefkowitz, E..
797 Amsterdam: Elsevier. pp. 735–741.

798 Prado, T., Silva, D.M., Guilayn, W.C., Rose, T.L., Gaspar, A.M.C., Miagostovich, M.P., 2011.
799 Quantification and molecular characterization of enteric viruses detected in effluents from
800 two hospital wastewater treatment plants. Water Res. 45, 1287–1297.
801 doi:10.1016/j.watres.2010.10.012

802 Pusch, D., Oh, D.-Y., Wolf, S., Dumke, R., Schröter-Bobsin, U., Höhne, M., Röske, I., Schreier, E.,
803 2005. Detection of enteric viruses and bacterial indicators in German environmental
804 waters. Arch. Virol. 150, 929–47. doi:10.1007/s00705-004-0467-8

805 Räsänen, S., Lappalainen, S., Kaikkonen, S., Hämäläinen, M., Salminen, M., Vesikari, T., 2010.
806 Mixed viral infections causing acute gastroenteritis in children in a waterborne outbreak.
807 Epidemiol. Infect. 138, 1227–1234. doi:10.1017/S0950268809991671

808 Rodriguez-Manzano, J., Alonso, J.L., Ferrús, M.A., Moreno, Y., Amorós, I., Calgua, B., Hundesa,
809 A., Guerrero-Latorre, L., Carratala, A., Rusiñol, M., Girones, R., 2012. Standard and new
810 faecal indicators and pathogens in sewage treatment plants, microbiological parameters
811 for improving the control of reclaimed water. Water Sci. Technol. 66, 2517–23.
812 doi:10.2166/wst.2012.233

813 Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009. Metagenomic analysis of viruses
814 in reclaimed water. Environ. Microbiol. 11, 2806–2820. doi:10.1111/j.1462-
815 2920.2009.01964.x

816 Rosario, K., Symonds, E.M., Sinigalliano, C., Stewart, J., Breitbart, M., 2009. Pepper Mild Mottle
817 Virus as an Indicator of Fecal Pollution. Appl. Environ. Microbiol. 75, 7261–7267.
818 doi:10.1128/AEM.00410-09

819 RUTJES, S.A., BOUWKNEGT, M., van der GIESSEN, J.W., de RODA HUSMAN, A.M., REUSKEN,
820 C.B.E.M., 2014. Seroprevalence of Hepatitis E Virus in Pigs from Different Farming Systems
821 in The Netherlands. J. Food Prot. 77, 640–642. doi:10.4315/0362-028X.JFP-13-302

822 Santiago-Rodriguez, T.M., Ly, M., Bonilla, N., Pride, D.T., 2015. The human urine virome in
823 association with urinary tract infections. Front. Microbiol. 6, 14.
824 doi:10.3389/fmicb.2015.00014

825 Sarma, R.H., Saram, M.H. (Mukti H.., State University of New York at Albany., 1990. Structure
826 & methods : proceedings of the Sixth Conversation in the Discipline Biomolecular
827 Stereodynamics held at the State University of New York at Albany, June 6-10, 1989.
828 Adenine Press.

829 Sauvage, V., Laperche, S., Cheval, J., Muth, E., Dubois, M., Boizeau, L., Hébert, C., Lionnet, F.,
830 Lefrère, J.-J., Eloit, M., 2016. Viral metagenomics applied to blood donors and recipients at
831 high risk for blood-borne infections. Blood Transfus. 14, 400–7. doi:10.2450/2016.0160-15

832 Schmieder, R., Edwards, R., 2011. Fast identification and removal of sequence contamination
833 from genomic and metagenomic datasets. PLoS One 6. doi:10.1371/journal.pone.0017288

834 Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors - occurrence, properties
835 and removal. J. Appl. Microbiol. 113, 1014–1026. doi:10.1111/j.1365-2672.2012.05384.x

836 Shinohara, T., Matsuda, M., Cheng, S.H., Marshall, J., Fujita, M., Nagashima, K., 1993. BK virus
837 infection of the human urinary tract. J. Med. Virol. 41, 301–5.

838 Smelov, V., Arroyo Mühr, L.S., Bzhalava, D., Brown, L.J., Komyakov, B., Dillner, J., 2014.
839 Metagenomic sequencing of expressed prostate secretions. J. Med. Virol. 86, 2042–2048.
840 doi:10.1002/jmv.23900

841 Smelov, V., Bzhalava, D., Arroyo Muhr, L.S., Eklund, C., Komyakov, B., Gorelov, a, Dillner, J.,
842 Hultin, E., 2016. Detection of DNA viruses in prostate cancer. Sci Rep 6, 25235.
843 doi:10.1038/srep25235

844 Spisák, S., Solymosi, N., Ittzés, P., Bodor, A., Kondor, D., Vattay, G., Barták, B.K., Sipos, F.,
845 Galamb, O., Tulassay, Z., Szállási, Z., Rasmussen, S., Sicheritz-Ponten, T., Brunak, S.,
846 Molnár, B., Csabai, I., 2013. Complete Genes May Pass from Food to Human Blood. PLoS

847 One 8, e69805. doi:10.1371/journal.pone.0069805

848 Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples
849 from the United States. Appl. Environ. Microbiol. 75, 1402–9. doi:10.1128/AEM.01899-08

850 Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., Kamagata, Y., Liu, W.-
851 T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical
852 climate. Environ. Microbiol. 14, 441–452. doi:10.1111/j.1462-2920.2011.02630.x

853 Tan, L. Van, van Doorn, H.R., Nghia, H.D.T., Chau, T.T.H., Tu, L.T.P., de Vries, M., Canuti, M.,
854 Deijs, M., Jebbink, M.F., Baker, S., Bryant, J.E., Tham, N.T., BKrong, N.T.T.C., Boni, M.F., Loi,
855 T.Q., Phuong, L.T., Verhoeven, J.T.P., Crusat, M., Jeeninga, R.E., Schultsz, C., Chau, N.V.V.,
856 Hien, T.T., van der Hoek, L., Farrar, J., de Jong, M.D., 2013. Identification of a new
857 cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections.
858 MBio 4. doi:10.1128/mBio.00231-13

859 Tate, J.E., Burton, A.H., Boschi-Pinto, C., Steele, A.D., Duque, J., Parashar, U.D., WHO-
860 coordinated Global Rotavirus Surveillance Network, 2012. 2008 estimate of worldwide
861 rotavirus-associated mortality in children younger than 5 years before the introduction of
862 universal rotavirus vaccination programmes: a systematic review and meta-analysis.
863 Lancet. Infect. Dis. 12, 136–41. doi:10.1016/S1473-3099(11)70253-5

864 Tomlinson, J.A., Faithfull, E., Flewett, T.H., Beards, G., 1982. Isolation of infective tomato bushy
865 stunt virus after passage through the human alimentary tract. Nature 300, 637–8.

866 Tseng, C.-H., Tsai, H.-J., Kirkwood, C.D., Wang, D., Hay, C., Hallenbeck, P., Knowles, N., Lemon, S.,
867 Minor, P., Pallansch, M., Palmenberg, A., Skern, T., 2007. Sequence analysis of a duck
868 picornavirus isolate indicates that it together with porcine enterovirus type 8 and simian
869 picornavirus type 2 should be assigned to a new picornavirus genus. Virus Res. 129, 104–
870 114. doi:10.1016/j.virusres.2007.06.023

871 Vasilyev, E. V, Trofimov, D.Y., Tonevitsky, A.G., Ilinsky, V. V, Korostin, D.O., Rebrikov, D. V, 2009.
872 Torque Teno Virus (TTV) distribution in healthy Russian population. Virol. J. 6, 134.
873 doi:10.1186/1743-422X-6-134

874 Vega, E., Barclay, L., Gregoricus, N., Williams, K., Lee, D., Vinjé, J., 2011. Novel surveillance
875 network for norovirus gastroenteritis outbreaks, United States. Emerg. Infect. Dis. 17,
876 1389–1395. doi:10.3201/eid1708.101837

877 Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R.,

878 Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi,
879 J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1.
880 doi:10.1371/journal.pbio.0000002

881 Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., Silva, J., Tammadoni, S., Nosrat,
882 B., Conrad, D., Rohwer, F., 2009. Metagenomic Analysis of Respiratory Tract DNA Viral
883 Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals. PLoS One 4, e7370.
884 doi:10.1371/journal.pone.0007370

885 Zhang, T., Breitbart, M., Lee, W.H., Run, J.-Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T.,
886 Rohwer, F., Ruan, Y., 2006. RNA viral community in human feces: prevalence of plant
887 pathogenic viruses. PLoS Biol. 4, e3. doi:10.1371/journal.pbio.0040003

888 Zhang, Y., Zhu, Z., Yang, W., Ren, J., Tan, X., Wang, Y., Mao, N., Xu, S., Zhu, S., Cui, A., Zhang, Y.,
889 Yan, D., Li, Q., Dong, X., Zhang, J., Zhao, Y., Wan, J., Feng, Z., Sun, J., Wang, S., Li, D., Xu, W.,
890 2010. An emerging recombinant human enterovirus 71 responsible for the 2008 outbreak
891 of hand foot and mouth disease in Fuyang city of China. Virol. J. 7, 94. doi:10.1186/1743-
892 422X-7-94

893 Zhou, N., Lv, D., Wang, S., Lin, X., Bi, Z., Wang, H., Wang, P., Zhang, H., Tao, Z., Hou, P., Song, Y.,
894 Xu, A., 2016. Continuous detection and genetic diversity of human rotavirus A in sewage in
895 eastern China, 2013–2014. Virol. J. 13, 153. doi:10.1186/s12985-016-0609-0

896

897

Table 1
Click here to download Table: Table_1_JFA_JRM.docx

Table 1. Potentially pathogenic human viral families detected in raw sewage by using 10L Skimmed Milk Flocculation (SMF).

Family	Genus/Species	Winter	Spring	Summer	Total hits
Astroviridae	Mamastrovirus 1	48	7	5	60
	Mamastrovirus 6	8	1	10	19
	Mamastrovirus 8/9	18	0	2	20
Parvoviridae	Human Bocavirus 2	6	1	0	8
	Human Bocavirus 3	3	0	0	3
	Human Bocavirus 4	1	2	0	3
	Adeno-associated virus 2	12	4	1	17
	Adeno-associated virus 3	2	0	0	2
	Adeno-associated virus 5	2	0	0	2
Caliciviridae	Human Sapovirus GI	7	2	4	13
	Human Sapovirus GII	11	0	3	14
	Norovirus GI	17	6	3	26
	Norovirus GII	20	1	1	22
Polyomaviridae	JC Polyomavirus	4	0	0	4
Circoviridae	Human circovirus	5	12	2	19
Picornaviridae	Hepatitis A virus	2	0	0	2
	Salivirus	20	20	22	62
	Cosavirus	0	2	1	3
	Rosavirus	1	0	0	1
	Enterovirus A	4	1	1	6
	Enterovirus B	10	4	5	19
	Enterovirus C	5	2	1	8
	Enterovirus D	1	0	0	1
	Enterovirus J	0	0	1	1
	Rabovirus	10	0	1	11
	Cardiovirus	8	7	5	20
	Aichi virus	16	12	42	70
Hepeviridae	Hepatitis E virus	4	0	0	4
Picobirnaviridae	Human picobirnavirus	37	11	15	63

Figure 1

Click here to download Figure: fig 1 mapa RAWS_JRM.docx

Figure 1. Sampling sites locations. Site A: Sant Adrià del Besòs WWTP; Site B: Granollers WWTP.

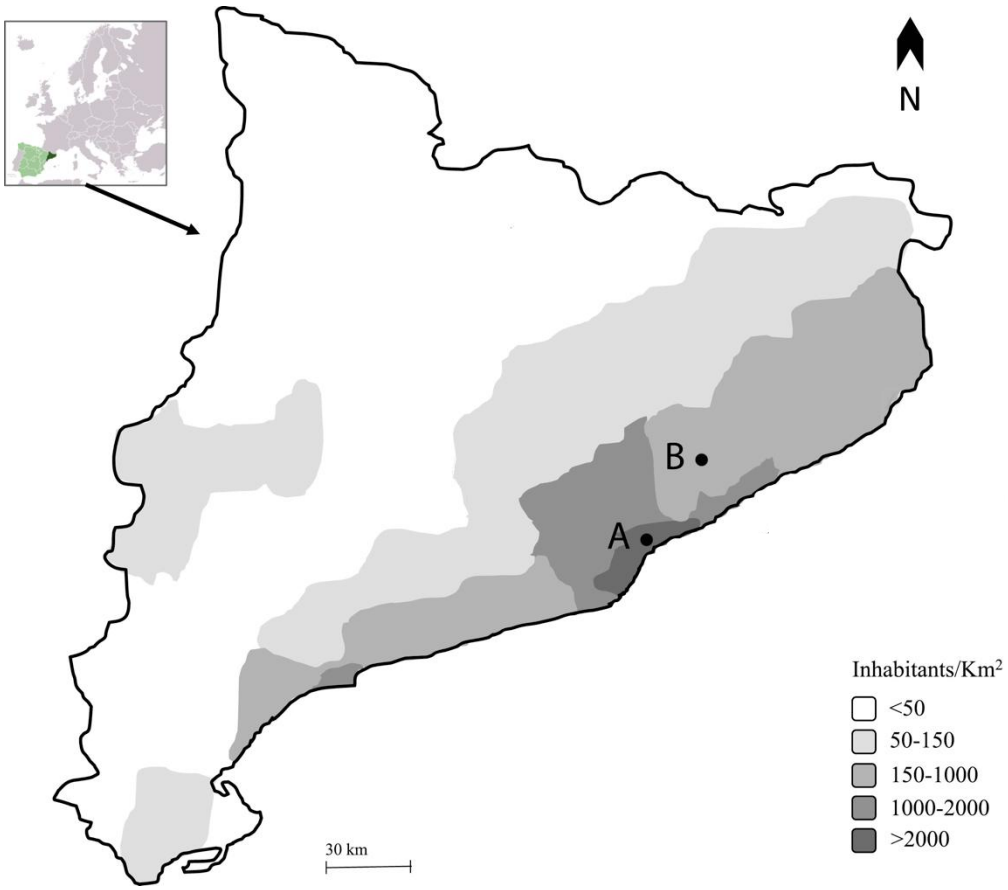

Figure 2
Click here to download Figure: Figure_2_JFA_JRM (1).docx

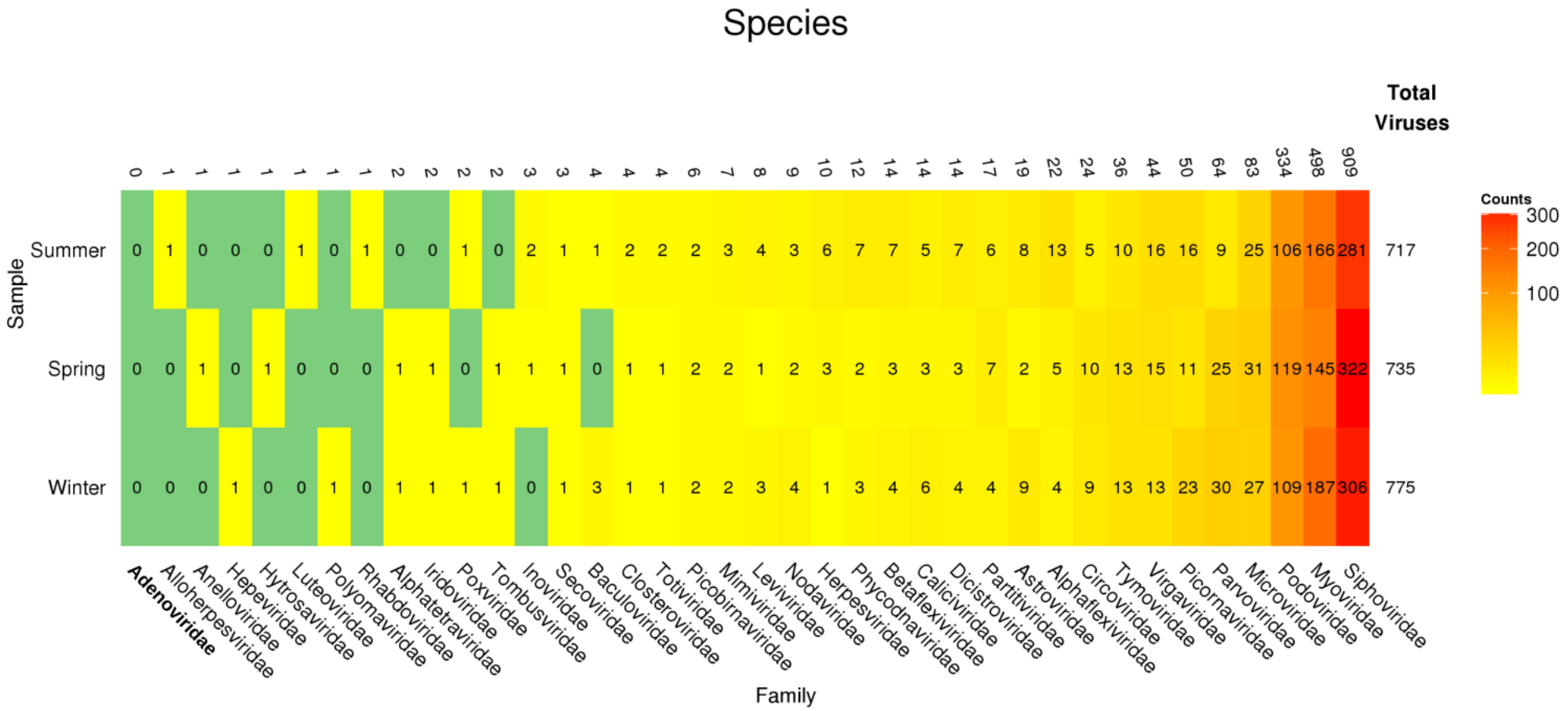

Figure 2. Heatmap profile showing the relative abundance of viral species detected over three different seasonal samples. Each cell contains the number of different species that had at least a sequence with a positive BLAST hit that passed all the selection criteria. The top row indicates the count sums of viral species classified by 37 families and the right column correspond to the count sums of viral species based on each seasonal sample. Data spanned from green (not detected) to red (high relative abundance), as illustrated by the colour scale.

Figure 3
Click here to download Figure: Figure_3_JFA_JRM.docx

Figure 3. Phylogenetic tree based on the complete nucleotide sequence of the Adenovirus hexon gene. Detected Adenovirus sequences from 454 sequencing experiment that presented a match in sewage were aligned. Human adenovirus species are shown in boldface, where colored arcs highlight the distinct taxon groups ranging from HAdV-A to HAdV-G. The phylogenetic tree was constructed using the neighbor-joining method and 1000 bootstrap replicates (bootstrap values are shown on the tree branches).

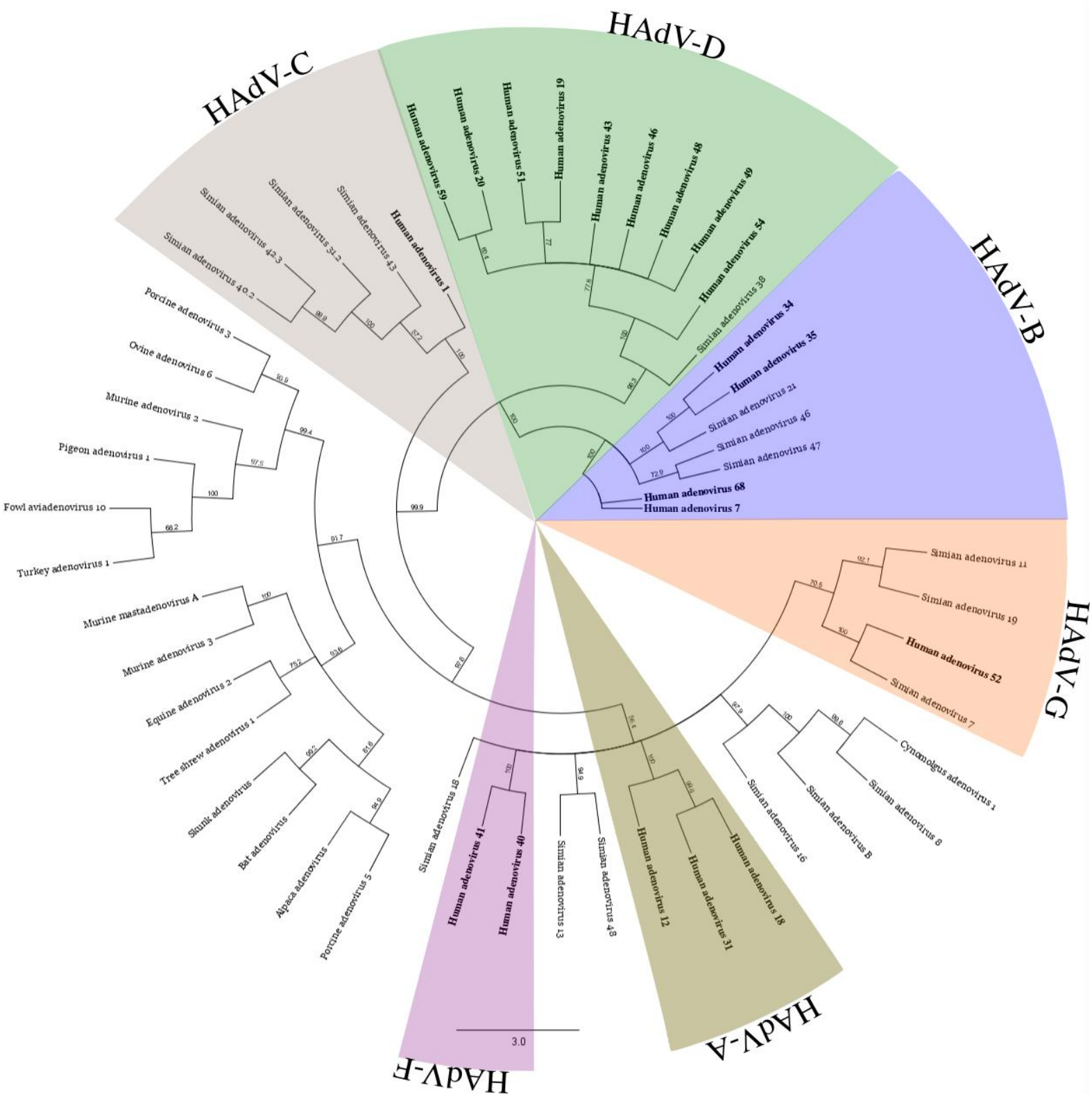

Figure 4
Click here to download Figure: Figure_4_JFA_JRM.docx

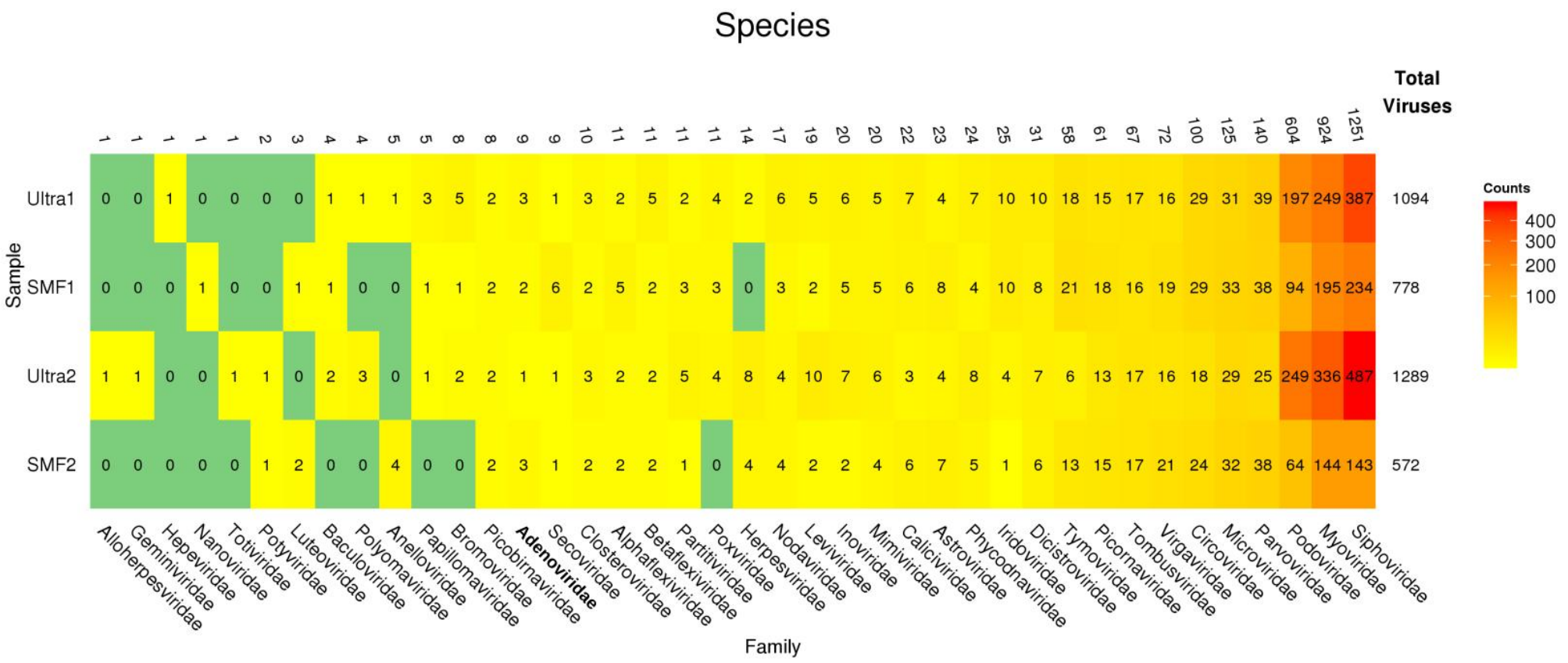

Figure 4.
Heatmap profile showing the relative abundance of viral species detected over four different concentration methods. Each cell contains the number of different species that had at least a sequence with a positive BLAST hit that passed all the selection criteria. The top row indicates the count sums of viral species classified by 40 families and the right column correspond to the count sums of viral species based on each sampling method. Data spanned from green (not detected) to red (high relative abundance), as illustrated by the colour scale.