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


a Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain
b Computational Genomics Lab, University of Barcelona and Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Catalonia, Spain
c Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom

HIGHLIGHTS

• Sensitive methods have been developed to study the metavirome of sewage
• The metavirome can pinpoint important pathogens circulating in the human population
• Mainly DNA viruses are excreted in urine, being JCPyV the most abundant
• The sewage metavirome is a useful tool for metagenomics in public health surveillance

ABSTRACT

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 quantitative PCR (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, 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
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; Pushch et al., 2005). This higher viral survival in STP treated water can represent a threat to consumers because FIB 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 through the human intestines. Sewage also contains other non-human inputs, which increase the diversity of this complex ecosystem. Viruses do not have conserved molecular markers, such as 16S rRNA, that are shared across all species hampering 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 were 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 an 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.

The infectivity of the known and unknown viral species present in raw sewage was explored by intravenously inoculating a sewage sample into rhesus macaques 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 flocculation

Sampling points of this study are presented in Fig. 1. Three 10 L samples of raw sewage from a STP in Sant Adria del Besòs 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 interferring 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/cm² was added to each sample. After 8 h of stirring, flocks were centrifuged at 8,000 × g for 40 min, and the pellet was suspended in 4 mL of phosphate buffer [vol/vol] (0.2 M Na2HPO4 and 0.2 M NaH2PO4). 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-adapted protocol from Calgua et al. (2008) and 42 mL for the ultracentrifugation protocol adapted from Pina et al. (1998b). 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 to 63 years old), although most of them were living in Barcelona. Individual urine samples were ultracentrifuged for 1 h at 90,000 × g at 4 °C. The obtained viral pellets were suspended in 300 μL of 1 × 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 10,000 × g for 1.5 min. 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′-GTTCAGTCGATACTGNNNNNNN′-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′-GTTCAGTCGATACTG′-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 2 × 250 bp and 2 × 300 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; Altschul et al., 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 98% coverage cut-off, each sequence was classified into its likely taxonomic group. Tables summarizing the number of sequences from the assembly matching each taxonomic unit were built. From these tables, richness ratios were calculated using the sequences from the assembly matching each taxonomic unit were included in the package, the non-parametric model Chaos1 was chosen, which was the model that provided the best results for the datasets. Heatmaps were generated using ggplot2 R graphics library (Kolde, 2015).

2.4. Targeted metagenomics for the characterisation of adenovirus

To detect and characterise all mastadenoviruses and other potential AdV 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 AdV 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), Stadenovirus (4) and Ichtadenovirus (1). OTUs that matched the 0.02 criteria were blasted against the AdV local database using BLASTN (Altschul et al., 1997; Altschul et al., 1990). A phylogenetic tree using Raxml with 1000 bootstrap replicates was computed using Geneious, version 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 10× 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 differentlibrary preparations were sequenced: a pooled library prior to the raw sewage inoculation from the two rhesus macaques (PW1), two different libraries from each of the animals one week after the inoculation (RW1 and RW2), and a pooled library from both rhesus macaques 4 weeks after the inoculation (RW4).

3. Results and discussion

3.1. MiSeq run outputs in 10 L sewage samples from 3 different seasons

The MiSeq 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 Fig. 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 Virgaviroidae 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 contig associat ed 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/klaassevirus, 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 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.

Please cite this article as: Fernandez-Cassi, X., et al., Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance, Sci Total Environ (2017), https://doi.org/10.1016/j.scitotenv.2017.08.249
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 $3.18 \times 10^4$ GC/L, $5.32 \times 10^5$ GC/L and $1.23 \times 10^5$ GC/L, respectively.

### 3.1.1. 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 3677 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 taxaons. The detected AdVs from raw sewage are shown in Fig. 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 marine adenovirus 2 (60%) and HAdV from species F, including HAdV-41 (29%) and HAdV-40 (0.7%). In total, 20 human adenoviruses 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 HAdV-40 and 41 (see Fig. 3). Therefore, the detected SAdVs could be variants of the closely related HAdV-40 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 GSJunior has an overall error rate of 0.18%, and it is also known that the distribution of errors in the sequences is not homogenous.

### Table 1

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

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 $3.18 \times 10^4$ GC/L, $5.32 \times 10^5$ GC/L and $1.23 \times 10^5$ GC/L, respectively. 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 3677 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 taxaons. The detected AdVs from raw sewage are shown in Fig. 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 marine adenovirus 2 (60%) and HAdV from species F, including HAdV-41 (29%) and HAdV-40 (0.7%). In total, 20 human adenoviruses 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 HAdV-40 and 41 (see Fig. 3). Therefore, the detected SAdVs could be variants of the closely related HAdV-40 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 using metagenomics: 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 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 hamper its applicability. A recent comparative study published in collaboration with Hjelmso et al. (2017) showed that the analysis of 10 L SMF, as described in Section 2.1.1, in

---

Please cite this article as: Fernandez-Cassi, X., et al., Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance, Sci Total Environ (2017), https://doi.org/10.1016/j.scitotenv.2017.08.249
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. (1998a) 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 MiSeq 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 species detected using ultracentrifugation (see Fig. 4). The four viral concentrates were analysed for HAdV by qPCR showing $8.14 \times 10^5$ GC/L, $1.23 \times 10^5$ GC/L, $2.19 \times 10^5$ GC/L, and $1.48 \times 10^5$ 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 Fig. 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, 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...
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 HAdV. 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 after the RT and Sequenase reactions. 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 were detected, or putatively considered, as pathogenic for humans. Several sequences resembling animal parvoviruses that infect dogs, rats, cattle, swine and several densovirus species were also 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, some of them infecting human, 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 densovirus species were also 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, some of them infecting human, 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 densovirus species were also 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, some of them infecting human, 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).
have been identified. Human bocavirus (HBoV) species HBoV1, 2, 3 and, 4 and human bufaviruses have been observed, yet the implications of these paroviruses 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 in some cases, they are associated with gastroenteritis (Rutjes et al., 2014).

Orthohepevirus, within the Hepeviridae family, is the genus of the species Orthohepevirus 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-Abraham et al., 2009). The finding in two samples 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, 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, Pepper Mild Mottle Virus (PMMV) has been recently connected to specific immune responses, fever, and abdominal pain in humans by 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, Leviriviridae 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 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. (2009b) 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, 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; Rosario et al., 2009a), 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 faecal 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. 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 concentration 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 HPV-129 and HPV-170, were identified; the HPVVs probably originated from epithelial desquamation during urination. The detection of HPVVs 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 γ-HPV (HPV-49, HPV-92, and HPV-96) and γ-HPV (HPV-121 and HPV-178) 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; Smelov et al., 2014). Although skin desquamation and excretion through faeces might be the main modes through which human papillomaviruses arrive at sewage, the excretion of specific papillomaviruses such as the skin-specific γ-HPV, 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 HPVVs 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 (Vasilev 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 macaque 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 macaques is an effective method to replicate the virus (Pina et al., 1998b). 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 macaques did not present JCPyV or HEV sequences in its skin on the studied dates (one week and one month after inoculation). The pooled sample from both rhesus macaques 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 1462 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).

4. 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 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 circular DNA 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.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2017.08.249.

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 (2014SGR914) and IPI 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 de Spain (www.mineco.gob.es). During the development of this study, Xavier Fernandez-Cassi was a fellow of the Catalan Government "AGAUR"