*Title Page

CHARACTERISATION OF THE SEWAGE VIROME: COMPARISON OF NGS TOOLS AND OCURRENCE OF SIGNIFICANT PATHOGENS

Sandra Martínez-Puchol[1], Marta Rusiñol[1,2], Xavier Fernández-Cassi[1], Natàlia Timoneda[1,3], Marta Itarte[1], Cristina Andrés[4], Andrés Antón[4], Josep F Abril[3], Rosina Girones[1,2], Sílvia Bofill-Mas[1,2].

1. Laboratory of Viruses Contaminants of Water and Food; Genetics, Microbiology & Statistics Dept.; Universitat de Barcelona. Barcelona, Catalonia, Spain.

2. The Water Research Institute (IdRA); Universitat de Barcelona. Barcelona, Catalonia, Spain.

3. Computational Genomics Lab; Genetics, Microbiology & Statistics Dept.; Universitat de Barcelona; Institut de Biomedicina (IBUB); Barcelona, Catalonia,
Spain

4. Respiratory Viruses Unit, Virology Section, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona. Barcelona (Spain).

Corresponding author: Sandra Martínez-Puchol (smartinezpuchol@ub.edu)

ABSTRACT

NGS techniques are excellent tools to monitor and identify viral pathogens circulating among the population with some limitations that need to be overcome, especially in complex matrices. Sewage contains a high amount of other microorganisms that could interfere when trying to sequence viruses for which random PCR amplifications are needed before NGS. The selection of appropriate NGS tools is important for reliable identification of viral diversity amongst the population.

We have compared different NGS methodologies (Untargeted Viral Metagenomics, Target Enrichment Sequencing and Amplicon Deep Sequencing) for the detection and characterization of viruses in urban sewage, focusing on three important human pathogens: papillomaviruses, adenoviruses and enteroviruses.

A full picture of excreted viruses was obtained by applying Untargeted Viral Metagenomics, which detected members of four different human viral families in addition to bacteriophages, plant viruses and viruses infecting other hosts. Target Enrichment Sequencing, using specific vertebrate viral probes, allowed the detection of up to eight families containing human viruses, with high variety of types within the families and with a high genome coverage.

By applying Amplicon Deep Sequencing, the diversity of enteroviruses, adenoviruses and papillomaviruses observed was higher than when applying the other two strategies and this technique allowed the subtyping of an enterovirus A71 C1 strain related to a brainstem encephalitis outbreak occurring at the same time in the sampling area.

From the data obtained, we concluded that the different strategies studied provided different levels of analysis: TES is the best strategy to obtain a broad picture of human viruses present in complex samples such as sewage. Other NGS strategies are useful for studying the virome of complex samples when also targeting viruses infecting plants, bacteria, invertebrates or fungi (Untargeted Viral Metagenomics) or when observing the variety within a sole viral family is the objective of the study (Amplicon Deep Sequencing).

KEYWORDS

Sewage virome, target enrichment sequencing, amplicon deep sequencing, papillomavirus, adenovirus, enterovirus.

1. INTRODUCTION

Viruses excreted by humans in faeces, urine and skin desquamation together with animal viruses and viruses infecting invertebrates, plants, bacteria and fungi constitute the sewage virome, a complex matrix that contains a large variety of pathogenic and commensal viruses and could give important information on persistent and acute infections affecting the population.

Currently, Next Generation Sequencing (NGS) methodologies are increasingly viewed as promising tools for the comprehensive study of microorganisms in a wide variety of samples and settings, replacing traditional molecular methods. The main reasons for this new trend are: the capacity to process larger number of samples simultaneously, the reduction in sequencing costs, the ease with which samples and libraries can be prepared and the development of faster and more

efficient bioinformatic tools with which to process the huge amount of data. (Huang et al., 2019).

The application of NGS techniques was easily incorporated into the study of bacterial communities even for environmental samples, but the lack of shared regions in viral genomes make them more challenging for viral sequencing. To overcome this problem, the use of random-primer-based sequencing facilitates the development of different protocols adapted for viral NGS (Cantalupo et al., 2011; Kohl et al., 2015; Ng et al., 2012), making this methodology key for the study of viral communities and viral discovery while providing enough starting material for library preparation.

The excreted virome of a population provides critical information about virus circulation, the introduction of emergent viruses and how they are transmitted among the population. The accurate study of the virome with a focus on specific target viral groups urges the development of NGS protocols with higher sensitivity.

The main limitation when analysing the virome from any type of sample is the low proportion of viral sequences identified compared with the total number of sequences amplified when using random primers (Krishnamurthy and Wang, 2017; Santiago-Rodriguez and Hollister, 2019; Tamaki et al., 2012). This is especially critical for human viruses which seem to represent a low fraction of the whole virome of sewage, clearly dominated by phages and plant viruses (Cantalupo et al., 2011; Ng et al., 2012). Few published studies in viral sewage metagenomics describe the human virome together with protocols to achieve a better recovery of viral sequences by applying mainly negative viral selection methods, also called pre-extraction, generally entail the use of, for example, filters, density gradients and nucleases (Cantalupo et al., 2011; Fernandez-Cassi et al.,

2018; Hjelmsø et al., 2017). Viral selection approaches are effective for avoiding the high background presence of other non-desired nucleic acids from bacteria or other hosts. Apart from negative selection methods, positive selection methods, also referred to post-extraction methods or Target Enrichment Methods, are characterised by the use of probes within the PCR assays, microarrays or hybridisation (Kumar et al., 2017). These methods increase in viral sequence reads as well as in the breadth and depth of genome coverage, in some cases extending to the full genome (Paskey et al., 2019; Wylie et al., 2015). Among these, the most commonly used are those based on hybridisation probes, with different custom-made and commercial approaches available (Chalkias et al., 2018; Hjelmsø et al., 2019; Mühlemann et al., 2018), showing the potential use in viral discovery (Briese et al., 2015).

Another NGS approach useful for identification and typing of different viral species for outbreak control or environmental surveillance is Amplicon Deep Sequencing, based on the mass sequencing of traditionally sanger-sequenced PCR amplicons. This approach opened up new opportunities promoting the detailed study of specific families and their diversity within a sample, also proving useful for a different nature of samples, from studies of quasispecies of Hepatitis C Virus or Human Immunodeficiency Virus in clinical settings (Del Campo et al., 2018; Kustin et al., 2019) to environmental samples (Fernandez-Cassi et al., 2018; Hata et al., 2018; Prevost et al., 2015).

This study shows the wide diversity of viral pathogens identified in urban sewage over one year of study with the application of different NGS tools for the determination of the virome with different levels of analysis. To our knowledge, this is the first work aimed to compare two different NGS strategies, Untargeted

Viral Metagenomics (UVM) and Target Enrichment Sequencing (TES), for the characterisation of the virome excreted in a population in terms of viral diversity, specificity and genome coverage. Also, Amplicon Deep Sequencing (ADS), was examined as an alternative strategy when a deeper study on a concrete viral family is needed. The work is focused on three specific viral families, human adenoviruses (HAdV), human papillomaviruses (HPV) and human enteroviruses (EV), since they are important pathogens and some of them are persistently excreted by the population. In addition, we analysed enterovirus nucleotide sequences obtained from clinical samples for evaluating the capacity of the studied NGS strategies to catch up viruses causing clinical disease in the population.

2. MATERIAL AND METHODS

2.1 Sampling, viral particles concentration and nucleic acid extraction

In April 2016, a 24-hour urban raw sewage composite sample was collected from a wastewater treatment plant (WWTP), located in the north of Barcelona city (WWTP A), that treats a population equivalent of up to 2.8 million and receives domestic and industrial waste from the sewerage system. This composite sample was analysed to compare TES, UVM and ADS.

Additionally, monthly samples were collected over the year, in 2016, from a second WWTP located 30 km from Barcelona (WWTP B). The samples from this WWTP were pooled by season (spring, summer, autumn and winter) and analysed by ADS in order to evaluate specific viral groups, in depth, over one year. This plant serves up to 115000 population equivalents.

Samples were collected in sterile containers and kept at 4ºC until concentrated within 24 hours. Viral particles from 42 ml of sewage from each sample were

concentrated by ultracentrifugation, as previously described (Bofill-Mas et al., 2006). DNase treatment (TurboDNAse, Ambion) and extraction of nucleic acids was performed as described previously (Fernandez-Cassi et al., 2018) using QIAmp RNA Viral Mini Kit (Qiagen).

2.2 Untargeted Viral Metagenomics and Target Enrichment Sequencing

2.2.1 Random tagging of nucleic acids and pre-amplification

Samples were prepared prior to the library construction following the protocol described (Fernandez-Cassi et al., 2018; Fernandez-Cassi et al., 2018). Briefly, in order to analyse both RNA and DNA viruses, NA were retrotranscribed with SuperScript III enzymes (Life Technologies) and a random nonamer primer. The second cDNA strand was obtained using Sequenase 2.0 (Thermo Fisher Scientific) and a PCR of 25 cycles was performed in order to obtain enough dsDNA for the next steps. These PCR products were purified and concentrated with the Zymo DNA Clean & Concentrate kit (Zymo Research) and quantified using Quibit 2.0 HSdsDNA kit (Life Technologies).

2.2.2 Library construction

For each sample, libraries were constructed in duplicate using KAPA HyperPlus Library Preparation Kit (KAPA Biosystems, Roche). An enzymatic fragmentation was performed in the previously obtained purification with a starting concentration ranging from 1 to 3 ng/µl following the manufacturer's instructions. After the fragmentation and the end-repair and A-tailing reaction, the adapter's ligation was performed. Using the KAPA Single-Indexed Adapter Kit (KAPA Biosystems, Roche), each sample was paired with the desired index. A post-ligation clean-up was followed with a double-sized size selection with the magnetic AMPure XP Beads (Beckman Coulter) to select those fragments of between 250

and 450 bp. Using an LM-PCR, sample libraries were amplified (seven cycles of amplification for the adapters used) and purified. The quality of the resulting libraries was assessed by using Agilent Bioanalyzer DNA 1000 assay (Agilent Technologies) and the concentration was measured using Qubit 2.0 (Life Technologies). Then, libraries were mixed together in two replicates, to obtain two pools containing 1 μg each. One pool was directly sequenced on an Illumina MiSeq 2x300bp platform and the other was captured later by using specific viral probes.

2.2.3 Capture of viral sequences by using the VirCapSeq-VERT Capture Panel

One of the pooled libraries was captured using the VirCapSeq-VERT Capture Panel (Roche). This panel consists of approximately two million probes, covering the genomes of 207 viral taxa known to infect vertebrates, thus enabling the detection of viral sequences in complex sample types (Briese et al., 2015). Using the HyperCap Target Enrichment Kit (Roche) and the HyperCap Bead Kit (Roche), the sample was prepared and then hybridised with the VirCapSeq-VERT probes at 47ºC for 20 hours. Immediately after the hybridisation, the captured multiplex DNA sample was recovered with the Capture Beads (HyperCap Bead Kit, Roche), using a magnetic particle collector, and cleaned. The DNA captured, still bead-bounded, was amplified using an LM-PCR. This post-capture PCR was purified and the concentration and quality were checked, as mentioned before. Captured DNA was sequenced in the same run as non-captured DNA in an Illumina MiSeq 2x300bp platform.

2.3 Amplicon Deep Sequencing

Specific nested PCR, previously proved to be adequate for typing purposes, for HAdV, HPV and EV amplification were performed as described (De Roda Husman et al., 1995; Fernandez-Cassi et al., 2018; Forslund et al., 2003; WHO Regional

Office for Europe, 2015) with incorporation of Illumina adapters in the nested primers. Amplicons were purified from agarose gel with the QIAquick Gel Extraction Kit (Qiagen) and then sequenced with an Illumina MiSeq 2x300bp platform.

2.4 Clinical Enterovirus A71 sampling

Most of 25 EV-A71-studied cases by Vall d'Hebron Hospital Respiratory Viruses Unit corresponded to rhombencephalitis, while the remaining patients had hand-foot-mouth disease (2), gastroenteritis (2), aseptic meningitis (1), and acute bronchitis (1). EV-A71 was amplified from upper and lower respiratory tract samples from suspicious patients by using CE-marked commercial multiplex real-time RT-PCR-based assay (Anyplex II RV16 assay, Seegene, Korea). Total nucleic acids were previously extracted using NucliSens easyMAG (bioMérieux, Marcy líEtoile, France) according to the manufacturer's instructions and kept frozen (-20C) until use. An additional real-time RT-PCR (Seegene, Korea) was carried out to improve the detection of all EV strains (Gimferrer et al., 2015).

The partial coding sequences of the viral protein 1 (VP1) from EV amplicons, obtained according to the protocol recommended by the World Health Organisation (WHO Regional Office for Europe, 2015), were used to construct a phylogenetic tree.

2.5 Bioinformatics

2.5.1 *UVM and TES bioinformatic processing and taxonomical assignment*

Pair-end FASTAQ files generated from the sequencing were analysed using Genome Detective web-based software

(https://www.genomedetective.com/)(Vilsker et al., 2018a). Briefly, low-quality reads and adapters were trimmed using `Trimmomatic` (Bolger et al., 2014), viral reads were selected using DIAMOND alignment method and non-viral sequences were discarded. Subsequently, viral reads were assembled with `metaSPAdes` (Nurk et al., 2017) and taxonomically classified with NCBI-BLASTX and NCBI-BLASTN against NCBI RefSeq viral database (Vilsker et al., 2018b; Wheeler et al., 2007), using only the contigs with 70% identity cut-off. Richness Chao1 ratio was calculated using the `Catchall` software, version 4.0 (Allen et al., 2013).

2.5.2 *Amplicon Deep Sequencing*

For the study of the amplicons, 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). The cleaned reads were clustered using the software `CDHIT` with default parameters (Huang et al., 2010); the output was queried for sequence similarity using NCBI-BLASTN against a reference genome species-specific custom database for HADV, HPV and EV (Wheeler et al., 2007). Based on the best BLAST results (95% coverage and 95% identity cut-off), each cluster was classified into its taxonomic group.

2.5.3 *Enterovirus phylogenetic analyses*

A phylogenetic tree based on enterovirus VP1 nucleotide sequences of the EV-A71 C1 strains obtained from sewage and clinical samples was constructed by a using neighbour-joining method with `Geneious` software version 11.0 (https://www.geneious.com) , where branches having a bootstrap value below 70% were discarded for the Figure. Coxackievirus A16 sequence (GenBank accession number KT327162) was used as the outgroup.

3. RESULTS AND DISCUSSION

3.1 Study of the sewage virome using Untargeted Viral Metagenomics and Target Enrichment Sequencing

An urban sewage composite sample obtained after 24 hours of collection in WWTP A was analysed using UVM and TES in parallel. A total of 6.07 million viral reads were obtained when applying TES and 727,784 when applying UVM. Read counts for each of the viral groups detected are presented in **Table 1**. The Chao1 diversity index showed, as expected, a higher richness when using UVM, increasing from 170 obtained using the TES approach, to 311 when using UVM.

Probe enrichment increased the detection of sequences from vertebrate viruses by a factor of 81.04% (299,650 sequences) of the total sequences obtained, compared with the 2.74% (3,549 sequences) obtained without applying enrichment. An exhaustive description of the total virome obtained using both methodologies is presented as **Supplementary material 1**.

A comparison of the viral species obtained via each methodology, based on host distribution, is shown in **Figure 1**. Not only was the number of sequences from viruses that infect vertebrates higher after the TES approach, but also the number of species within each viral family increased. The *Picornaviridae* and *Parvoviridae* families showed a higher number of taxonomically assigned sequences after the application of the TES. When applying UVM, a wide variety of bacteriophage species was observed (49% of the total reads) within the families *Siphoviridae, Myoviridae, Microviridae* and *Podoviridae*, as well as species of the viral plant families (30% of the total reads) *Virgaviridae*, *Tombusviridae* and *Solemoviridae*. The predominance of these viral families in environmental settings has been reported before (Fernandez-Cassi et al., 2018) and is of importance because of

249 their potential economic impact as plant pathogens. Reads from vertebrate
250 families were identified as JC Polyomavirus, Mamastrovirus 1, Aichivirus A, non-
251 human circoviruses and parvoviruses, representing 2.75% of the total sequences.
252 Accordingly, UVM can be considered a reference tool for the study of a global
253 picture of the whole virome of environmental and clinical samples, in accordance
254 with other studies performed by our research group (Cantalupo et al., 2011; Xavier
255 Fernandez-Cassi et al., 2018).
256 The TES kit used in this study was developed in 2015 by Briese *et al*. (Briese et al.,
257 2015) with the aim of capturing viral sequences from only those viruses that infect
258 vertebrates, for use in clinical and veterinarian settings. This system uses more
259 than two million probes of 207 different vertebrate viruses and has been reported
260 useful for the detection of new variants. When applied to raw sewage, probes
261 helped to capture vertebrate viruses (Briese et al., 2015), although some
262 sequences from other viral hosts were still recovered. These kind of enrichment
263 approaches are clearly biased towards specific viral families (Parras-Moltó et al.,
264 2018) and while they might be useful for studying specific viral families (or groups
265 of families), they do not provide the whole virome picture.
266 However, when focusing on viral families that infect humans, TES provided a more
267 sensitive approach, allowing the detection of more viral families and catching a
268 wide diversity of viral species within a given family while providing a higher
269 number of reads. Members of the *Adenoviridae*, *Hepeviridae*, *Papillomaviridae*, and
270 *Reoviridae* families as well as some species of the *Picornaviridae* (including an
271 EV71 contig) and *Polyomaviridae* were only detected when using TES. Regarding
272 the *Astroviridae* and *Caliciviridae* families, a big increase in sequences from
273 Mamastrovirus 1 species (up to 3-log) and Norwalk and Sapporo viruses (up to 2-

log) was observed when compared with UVM. Again, when processing these reads as contigs, a considerable increase in the genome coverage of each virus was observed after TES which is of huge interest for viral characterisation and discovery, as reported before (Briese et al., 2015). Contigs corresponding to Aichivirus A, Norwalk virus, JC Polyomavirus and Mamastrovirus 1 presented more than 90% genome coverage. Indeed, for the latter two, almost the totality of the genome was covered, at 93.82% and 99.25% respectively, whereas without enrichment the respective coverage of these viruses was 77.37% and 30.45% **(Table 1).** This fact could be due to the higher number of sequences obtained with TES and the wider distribution of reads from across all viral genomes.

From the data obtained, we conclude that TES is the best strategy to obtain a broad picture of human viruses present in complex samples such as sewage, as well as being a technique that might be useful for environmental public health surveillance.

3.2 Application of Amplicon Deep Sequencing to the study of specific viral groups in sewage: human adenoviruses, papillomaviruses, and enteroviruses

3.2.1. *Comparison of ADS with UVM and TES strategies in the study of the sewage virome*

Amplicon Deep Sequencing has demonstrated its usefulness in the study of specific viral groups, giving higher sensitivity and a large amount of information about the viruses present in a given sample. This methodology has previously been applied to environmental samples for studying diversity within viral groups, such as *Adenoviridae* (Fernandez-Cassi et al., 2018; Ogorzaly et al., 2015), *Papillomaviridae*

(Hamza and Hamza, 2018; Iaconelli et al., 2015; La Rosa et al., 2015) or enterovirus (Brinkman et al., 2017).

The composite 24h sewage sample collected from WWTP A was evaluated by ADS using specific PCR primers for HAdV, HPV and EV families. Results obtained after the application of ADS are summarised in **Figure 2**.

While ADS showed similar adenovirus species to TES, detecting Human Mastadenovirus F and A species, with HAdV 41 and HAdV 18 being the most prevalent, some of the species appeared only with TES (e.g., HAdV5, HAdV 27 and HAdV 56) and others only with ADS (e.g., HAdV 18, HAdV 12 and HAdV 61).

Regarding HPV, ADS provided a higher diversity of species than TES, with HPV6 and HPV66 being the only alphapapillomavirus detected by ADS and the most abundant types, being 36.01% and 21.22% of the total sequences detected respectively. Other members of the betapapillomavirus genus were also detected, with HPV120 and HPV19 being the most prevalent types. An important difference between these two methodologies was observed: sequences of HPV 122 and HPV 49 (betapapillomavirus types) were obtained from both approaches, but HPV 17 (betapapillomavirus type) and HPV127 (gammapapillomavirus type) only by TES and a wider variety of types only by ADS, including oncogenic ones.

Regarding enterovirus, viral species Enterovirus A, B and C were detected by ADS as occurred when applying TES. E-E30 was the most prevalent enterovirus, followed by EV-C99 and EV-A119. Also, sequences from EV-A71 were obtained by ADS and TES, which were of interest because a clinical brainstem encephalitis outbreak was occurring in Catalonia during the sampling period. By applying TES, only one EV-A71 contig (884 bp) was obtained from WWTP A and, by ADS, 14 amplicons (301 bp) were obtained from both WWTPs.

3.2.2. *Amplicon Deep Sequencing for characterization of specific viral groups in sewage: human adenoviruses, papillomaviruses, and enteroviruses*

The distribution of the three viral groups over one year in sewage was also evaluated by applying ADS in seasonally pooled samples from WWTP B and the results obtained are represented in **Figure 3**. A wide variety of HAdV, HPV and EV sequences was obtained.

The in-depth analysis of HAdV showed more than 10 serotypes in the year of study **(Figure 3a)**. HAdV from *Human Mastadenovirus* A species, related to gastrointestinal, urinary and respiratory infections, and *Human Mastadenovirus* F, related to infantile gastroenteritis, were the groups most commonly detected over the whole year, with being HAdV 41 (Group F) and HAdDV 31 (Group A) being the most abundant types. HAdV 40 (Group F), and HAdV 12 (Group A) found throughout the year and HAdV 18 (Group A) in spring. Other *Human Mastadenoviruses* serotypes (HAdV 51, HAdV 59, HAdV 46, HAdV 19) were detected in lower proportions, as previously described in other studies (Fernandez-Cassi et al., 2018; Iaconelli et al., 2017).

Many HAdV produce infections that may be subclinical and could be excreted in faeces by healthy people over a long period. The prevalence of HAdV in sewage through the year is in accordance with the proposal to use adenovirus as an indicator of human faecal contamination (Pina et al., 1998).

For papillomaviruses, a large diversity of types was observed by ADS through the whole year of sampling **(Figure 3b)**. Sequences of the potentially oncogenic types HPV 6 and HPV 66 (genus alphapapillomavirus) were detected through the year, with higher proportions than other HPV in the spring and winter months. Members of this genus had been described previously in raw sewage (Iaconelli et al., 2015), but HPV 66, involved in vulvar cancer and classified as a group I carcinogen (Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Biological Agents., 2009), has never been reported in environmental samples. More than 25 different cutaneous betapapillomaviruses were detected, with HPV 120, HPV 19, HPV 9, HPV 8, HPV 49 and HPV 80 being the most abundant during the year of observation.

Enterovirus ADS showed different distribution patterns through the studied year **(Figure 3c)**: CV-B5, E-E18 and E30 (EV species B) were the serotypes more frequently detected in spring and summer and E-E14 (EV species B) and EV-C99 (EV species C) in autumn and winter. These enterovirus species are traditionally related to aseptic meningitis cases and, with the exception of Enterovirus C99, had been also reported as etiological agents involved in Acute Flaccid Paralysis (AFP) cases by the Spanish Ministry of Health in their Annual Epidemiologic Surveillance (López-Perea et al., 2017). This emergent member of enterovirus species C has been reported to be involved in the development of AFP (Brown et al., 2009) and was recently described as being very prevalent in Uruguay and Brazil (Lizasoain et al., 2018; Luchs et al., 2019).

Also, a comparison of EV presence in spring in two different WWTPs (composite sample from WWTP A, which was collected in spring, and spring pooled samples from WWTP B) showed E-E30 to be the most prevalent strain detected in both

374 plants while CV-B5 was the second most frequent strain in plant B and EV-C99 and
375 EV-119 in plant A.
376 In fact, the comparison of the results obtained by analysing sewage samples
377 collected in spring from both WWTPs by ADP resulted in a similar distribution of
378 HAdV, HPV and EV, despite the locations being 30 km apart and treating sewage
379 from different population equivalent amounts.
380 Further research should be directed towards analysing sewage during longer
381 periods, in order to establish the most appropriate sampling design for use in the
382 characterisation of the sewage virome from a determined population (composite,
383 pooled or individual samples), as well as for elucidating if there exist any
384 seasonality in the excretion of relevant pathogenic viruses.
385

386 **3.3 Effect of the EV-A71 outbreak on the excreted virome in sewage**

387 An outbreak of brainstem encephalitis occurred in the geographical area studied
388 during the studied period affecting more than 100 children showing a peak in
389 spring 2016 (Casas-Alba et al., 2017).
390 Sequences from EV-A71 were detected using ADS in both of the WWTP studied
391 which belong to geographical areas located within 30Km in the outbreak area. A
392 total of 14 amplicons of 301bp from VP1 region were subtyped as C1 with the
393 Enterovirus Genotyping tool (Kroneman et al., 2011). These sequences showed a
394 pairwise identity of 97% with sequences related to the German outbreak of 2015
395 (Böttcher et al., 2016) and with clinical sequences from patients involved in the
396 outbreak, obtained at Hospital Vall d'Hebron in Barcelona (Andres et al., 2019).
397 A phylogenetic tree including all these sequences was constructed and
398 represented in **Figure 4** and, even though the presence of this serotype was in a

minority, these sequences proved to be phylogenetically close to those that caused the German (2015) and Catalan outbreaks (2016).

By applying TES, only one EV-A71 contig (884 bp) was obtained from WWTP A, but subtyping was not possible due to the short VP1 region present in this contig.

Thus, the application of ADS, in addition to TES, has also been useful to monitor an outbreak, detecting EV A71 sequences only when an encephalitis outbreak was occurring simultaneously with the sampling period in two different sewage samples from two different locations in Catalonia.

4. CONCLUSIONS

- Enterovirus A71 C1 was detected by Amplicon Deep Sequencing and Target Enrichment Sequencing during an encephalitis outbreak, although it was only a small percentage of the enterovirus excreted, being other serotypes much more abundant. Only Amplicon Deep Sequencing was useful for subtyping purposes.
- Sewage from two different WWTPs, collected as a unique composite and as monthly pooled samples in spring 2016, showed a similar distribution of HAdV, HPV and EV types, despite the locations being 30 km apart and treating sewage from different population equivalent amounts.
- Untargeted Viral Metagenomics is the only NGS technique that provides a complete picture of the whole virome present in sewage including vertebrate, invertebrate, bacteria, plant and fungi viruses.
- Target Enrichment NGS based on probe capture has proved a very successful strategy for the study of vertebrate viruses in sewage samples

providing a higher number of detected families, a higher number of members within these families, more reads and larger genome coverage than conventional Untargeted Viral Metagenomics.

- Amplicon Deep Sequencing proved useful when observing the variety within a sole viral family is the objective of the study and, because is a very sensitive technique, it may be useful for the surveillance of specific pathogenic viruses (e.g: EV-A71).

ACKNOWLEDGEMENTS

We thank the Genomics Core Facility (UPF) for providing support in the sequencing experiments. We thank Eva Forés, Ayalkibet Hundesa, David Aguado and Aiora Aregita for their lab assistance and support. Graphical Abstract was created with Biorender.com.

AVAILABILITY OF DATA AND MATERIAL

The datasets generated during the current study are available in zenodo under the DOI number 10.5281/zenodo.3539112.

FUNDING

This work was supported through a grants funded by the Spanish Ministry of Economy and Competitiveness (MINECO) AGL2014-55081-R, AGL2017-86797-C2-1-R and JPI Water project METAWATER (4193-00001B). During the development of this study, Sandra Martínez-Puchol was a fellow of the MINECO. Sílvia Bofill-Mas is a Serra-Hunter fellow at the University of Barcelona.

REFERENCES

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. https://doi.org/10.1186/2049-2618-1-5

Andres, C., Vila, J., Gimferrer, L., Pinana, M., Esperalba, J., Codina, M.G., Barnes, M., Martin, M.C., Fuentes, F., Rubio, S., Alcubilla, P., Rodrigo, C., Pumarola, T., Anton, A., 2019. Surveillance of enteroviruses from paediatric patients attended at a tertiary hospital in Catalonia from 2014 to 2017. J. Clin. Virol. 110, 29–35. https://doi.org/10.1016/j.jcv.2018.11.004

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, 7894–7896. https://doi.org/10.1128/AEM.00965-06

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

Böttcher, S., Obermeier, P.E., Neubauer, K., Diedrich, S., Hofmann, J., Heider, H., Reif, U., Korn, K., Rabenau, H.F., Wille, A., Baillot, A., Pettke, A., Kühn, J., Ackermann, N., Ramelow, C., Hahn, R., Enders, M., Adam, M., Michel, D., Schubert, A., Weißbrich, B., Prifert, C., 2016. Recombinant enterovirus A71 subgenogroup C1 strains, Germany, 2015. Emerg. Infect. Dis. 22, 1843–1846. https://doi.org/10.3201/eid2210.160357

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., OJ, J., Lipkin, W., 2015. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. MBio 6, 1–11.

https://doi.org/10.1128/mBio.01491-15.Editor

Brinkman, N.E., Fout, G.S., Keely, S.P., 2017. Retrospective Surveillance of Wastewater To Examine Seasonal Dynamics of Enterovirus Infections. mSphere 2. https://doi.org/10.1128/msphere.00099-17

Brown, B.A., Maher, K., Flemister, M.R., Naraghi-Arani, P., Uddin, M., Oberste, M.S., Pallansch, M.A., 2009. Resolving ambiguities in genetic typing of human enterovirus species C clinical isolates and identification of enterovirus 96, 99 and 102. J. Gen. Virol. 90, 1713–1723. https://doi.org/10.1099/vir.0.008540-0

Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2, e00180-11. https://doi.org/10.1128/mBio.00180-11

Casas-Alba, D., de Sevilla, M.F., Valero-Rello, A., Fortuny, C., García-García, J.J., Ortez, C., Muchart, J., Armangué, T., Jordan, I., Luaces, C., Barrabeig, I., González-Sanz, R., Cabrerizo, M., Muñoz-Almagro, C., Launes, C., 2017. Outbreak of brainstem encephalitis associated with enterovirus-A71 in Catalonia, Spain (2016): a clinical observational study in a children's reference centre in Catalonia. Clin. Microbiol. Infect. https://doi.org/10.1016/j.cmi.2017.03.016

Chalkias, S., Gorham, J.M., Mazaika, E., Parfenov, M., Dang, X., DePalma, S., McKean, D., Seidman, C.E., Seidman, J.G., Koralnik, I.J., 2018. ViroFind: A novel target-enrichment deep-sequencing platform reveals a complex JC virus population in the brain of PML patients. PLoS One 13, e0186945. https://doi.org/10.1371/journal.pone.0186945

De Roda Husman, A.M., Walboomers, J.M.M., Van den Brule, A.J.C., Meijer, C.J.L.M., Snijders, P.J.F., 1995. The use of general primers GP5 and GP6

496 elongated at their 3’ ends with adjacent highly conserved sequences improves
497 human papillomavirus detection by PCR. J. Gen. Virol. 76, 1057–1062.
498 https://doi.org/10.1099/0022-1317-76-4-1057
499 Del Campo, J.A., Parra-Sánchez, M., Figueruela, B., García-Rey, S., Quer, J., Gregori,
500 J., Bernal, S., Grande, L., Palomares, J.C., Romero-Gómez, M., 2018. Hepatitis C
501 virus deep sequencing for sub-genotype identification in mixed infections: A real-
502 life experience. Int. J. Infect. Dis. 67, 114–117.
503 https://doi.org/10.1016/j.ijid.2017.12.016
504 Fernandez-Cassi, Xavier, Rusiñol, M., Martínez-Puchol, S., 2018. Viral concentration
505 and amplification from human serum samples prior to application of next-
506 generation sequencing analysis, in: Methods in Molecular Biology. pp. 173–188.
507 https://doi.org/10.1007/978-1-4939-8682-8_13
508 Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-
509 Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018.
510 Metagenomics for the study of viruses in urban sewage as a tool for public health
511 surveillance. Sci. Total Environ. 618, 870–880.
512 https://doi.org/10.1016/j.scitotenv.2017.08.249
513 Forslund, O., Ly, H., Higgins, G., 2003. Improved detection of cutaneous human
514 papillomavirus DNA by single tube nested ‘hanging droplet’ PCR. J. Virol.
515 Methods 110, 129–136. https://doi.org/https://doi.org/10.1016/S0166-
516 0934(03)00109-5
517 Gimferrer, L., Campins, M., Codina, M.G., Esperalba, J., Martin, M.D.C., Fuentes, F.,
518 Pumarola, T., Anton, A., 2015. First Enterovirus D68 (EV-D68) cases detected in
519 hospitalised patients in a tertiary care university hospital in Spain, October 2014.
520 Enferm. Infecc. Microbiol. Clin. 33, 585–589.

https://doi.org/10.1016/j.eimc.2015.01.008

Hamza, H., Hamza, I.A., 2018. Oncogenic papillomavirus and polyomavirus in urban sewage in Egypt. Sci. Total Environ. 610–611, 1413–1420. https://doi.org/10.1016/j.scitotenv.2017.08.218

Hata, A., Kitajima, M., Haramoto, E., Lee, S., Ihara, M., Gerba, C.P., Tanaka, H., 2018. Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters. Sci. Rep. 8, 11837. https://doi.org/10.1038/s41598-018-30217-y

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

Hjelmsø, M.H., Mollerup, S., Jensen, R.H., Pietroni, C., Lukjancenko, O., Schultz, A.C., Aarestrup, F.M., Hansen, A.J., 2019. Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. PLoS One 14, e0210368. https://doi.org/10.1371/journal.pone.0210368

Huang, B., Jennsion, A., Whiley, D., McMahon, J., Hewitson, G., Graham, R., De Jong, A., Warrilow, D., 2019. Illumina sequencing of clinical samples for virus detection in a public health laboratory. Sci. Rep. 9, 5409. https://doi.org/10.1038/s41598-019-41830-w

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003

Iaconelli, M., Petricca, S., Libera, S. Della, Di Bonito, P., La Rosa, G., 2015. First Detection of Human Papillomaviruses and Human Polyomaviruses in River Waters in Italy. Food Environ. Virol. https://doi.org/10.1007/s12560-015-9203-7

Iaconelli, M., Valdazo-Gonz Alez, B., Equestre, M., Ciccaglione, A.R., Marcantonio, C., Libera, S. Della, Rosa, G. La, 2017. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. https://doi.org/10.1016/j.watres.2017.05.039

Kohl, C., Brinkmann, A., Dabrowski, P.W., Radonic, A., Nitsche, A., Kurth, A., 2015. Protocol for Metagenomic Virus Detection in Clinical Specimens. Emerg. Infect. Dis. 21, 48–57. https://doi.org/10.3201/eid2101.140766

Krishnamurthy, S.R., Wang, D., 2017. Origins and challenges of viral dark matter. Virus Res. 239, 136–142. https://doi.org/10.1016/j.virusres.2017.02.002

Kroneman, A., Vennema, H., Deforche, K., v d Avoort, H., Penaranda, S., Oberste, M.S., Vinje, J., Koopmans, M., 2011. An automated genotyping tool for enteroviruses and noroviruses. J. Clin. Virol. 51, 121–125. https://doi.org/10.1016/j.jcv.2011.03.006

Kumar, A., Murthy, S., Kapoor, A., 2017. Evolution of selective-sequencing approaches for virus discovery and virome analysis. Virus Res. 239, 172–179. https://doi.org/10.1016/j.virusres.2017.06.005

Kustin, T., Ling, G., Sharabi, S., Ram, D., Friedman, N., Zuckerman, N., Bucris, E.D., Glatman-Freedman, A., Stern, A., Mandelboim, M., 2019. A method to identify respiratory virus infections in clinical samples using next-generation sequencing. Sci. Rep. 9, 2606. https://doi.org/10.1038/s41598-018-37483-w

La Rosa, G., Della Libera, S., Petricca, S., Iaconelli, M., Briancesco, R., Paradiso, R., Semproni, M., Di Bonito, P., Bonadonna, L., 2015. First detection of

Papillomaviruses and Polyomaviruses in swimming pool waters: unrecognized recreational water-related pathogens? J. Appl. Microbiol. 1–9. https://doi.org/10.1111/jam.12925

Lizasoain, A., Burlandy, F.M., Victoria, M., Tort, L.F.L., da Silva, E.E., Colina, R., 2018. An Environmental Surveillance in Uruguay Reveals the Presence of Highly Divergent Types of Human Enterovirus Species C and a High Frequency of Species A and B Types. Food Environ. Virol. 10, 343–352. https://doi.org/10.1007/s12560-018-9351-7

López-Perea, N., Masa-Calles, J., Torres de Mier, M., Fernandez-Martinez, B., Cabrerizo, M., 2017. Vigilancia de la Parálisis Flácida Aguda y Vigilancia de Enterovirus, 2016 Plan de Acción en España para la erradicación de la poliomelitis. https://doi.org/10.13140/RG.2.2.35292.90240

Luchs, A., Leal, E., Tardy, K., Milagres, F.A. de P., Komninakis, S.V., Brustulin, R., Teles, M. da A.R., Lobato, M.C.A.B.S., das Chagas, R.T., Abrão, M. de F.N.D.S., Soares, C.V. de D.A., Deng, X., Delwart, E., Sabino, E.C., da Costa, A.C., 2019. The rare enterovirus c99 and echovirus 29 strains in Brazil: potential risks associated to silent circulation. Mem. Inst. Oswaldo Cruz 114, e190160–e190160. https://doi.org/10.1590/0074-02760190160

Mühlemann, B., Jones, T.C., Damgaard, P. de B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smrčka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., Price, T.D., Moiseyev, V., Hansen, A.J., Orlando, L., Rasmussen, S., Sikora, M., Vinner, L., Osterhaus, A.D.M.E., Smith, D.J., Glebe, D., Fouchier, R.A.M., Drosten, C., Sjögren, K.-G., Kristiansen, K., Willerslev, E., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418–

423. https://doi.org/10.1038/s41586-018-0097-z

Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175. https://doi.org/10.1128/JVI.00869-12

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116

Ogorzaly, L., Walczak, C., Galloux, M., Etienne, S., Gassilloud, B., Cauchie, H.-M., 2015. Human Adenovirus Diversity in Water Samples Using a Next-Generation Amplicon Sequencing Approach. Food Environ. Virol. 7, 112–121. https://doi.org/10.1007/s12560-015-9194-4

Parras-Moltó, M., Rodríguez-Galet, A., Suárez-Rodríguez, P., López-Bueno, A., 2018. Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. Microbiome 6, 1–18. https://doi.org/10.1186/s40168-018-0507-3

Paskey, A.C., Frey, K.G., Schroth, G., Gross, S., Hamilton, T., Bishop-Lilly, K.A., 2019. Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. BMC Genomics 20, 155. https://doi.org/10.1186/s12864-019-5543-2

Pina, S., Puig, M., Lucena, F., Jofre, J., Girones, R., 1998. Viral pollution in the environment and in shellfish: Human adenovirus detection by PCR as an index of human viruses. Appl. Environ. Microbiol. 64, 3376–3382.

Prevost, B., Lucas, F.S., Ambert-Balay, K., Pothier, P., Moulin, L., Wurtzer, S., 2015.

621 Deciphering the Diversities of Astroviruses and Noroviruses in Wastewater
622 Treatment Plant Effluents by a High-Throughput Sequencing Method. Appl.
623 Environ. Microbiol. 81, 7215–7222. https://doi.org/10.1128/AEM.02076-15
624 Proceedings of the IARC Working Group on the Evaluation of Carcino- genic Risks to
625 Humans. Biological Agents., 2009. Proceedings of the IARC Working Group on
626 the Evaluation of Carcinogenic Risks to Humans. Biological Agents. IARC
627 Monogr. Eval. Carcinog. Risks to Humans 100B.
628 Santiago-Rodriguez, T.M., Hollister, E.B., 2019. Human virome and disease: High-
629 throughput sequencing for virus discovery, identification of phage-bacteria
630 dysbiosis and development of therapeutic approaches with emphasis on the human
631 gut. Viruses. https://doi.org/10.3390/v11070656
632 Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T.,
633 Kamagata, Y., Liu, W.-T., 2012. Metagenomic analysis of DNA viruses in a
634 wastewater treatment plant in tropical climate. Environ. Microbiol. 14, 441–452.
635 https://doi.org/10.1111/j.1462-2920.2011.02630.x
636 Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R.,
637 Alcantara, L.C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., de Oliveira,
638 T., 2018a. Genome Detective: An Automated System for Virus Identification from
639 High-throughput sequencing data. Bioinformatics.
640 https://doi.org/10.1093/bioinformatics/bty695
641 Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R.,
642 Alcantara, L.C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., de Oliveira,
643 T., 2018b. Genome Detective: an automated system for virus identification from
644 high-throughput sequencing data. Bioinformatics 1–3.
645 https://doi.org/10.1093/bioinformatics/bty695

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., Yaschenko, E., 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D5–D12. https://doi.org/10.1093/nar/gkl1031

WHO Regional Office for Europe, 2015. Enterovirus surveillance guidelines. Guidel. enterovirus Surveill. Support Polio Erad. Initiat.

Wylie, T.N., Wylie, K.M., Herter, B.N., Storch, G.A., 2015. Enhanced virome sequencing using targeted sequence capture. Genome Res. 25, 1910–1920. https://doi.org/10.1101/gr.191049.115

Table 1

Click here to download Table: TABLE 1 R1.docx

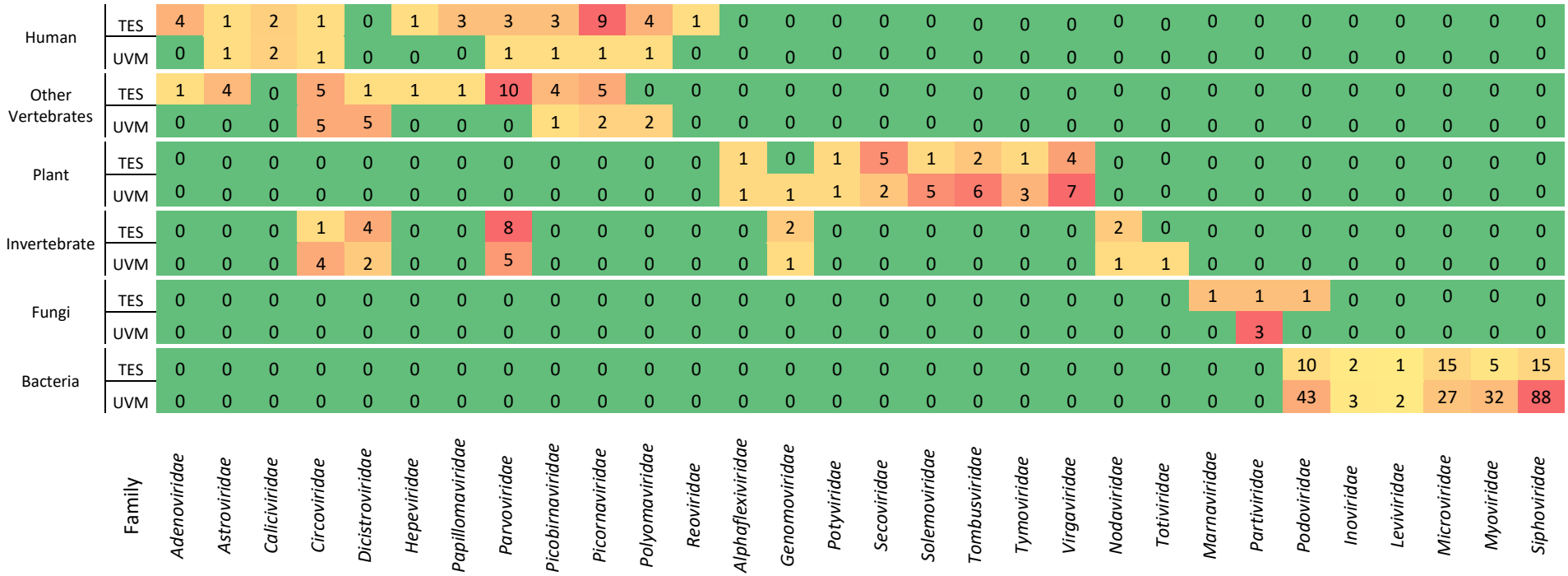
Table 1. Read counts and Genome coverage obtained from contigs of viral species of interest obtained by Target Enrichment Sequencing (TES) and Untargeted Viral Metagenomics (UVM).

		Viral sequence reads				
		Viral read count			Breadth genome coverage (%)	
Viral families	Viral species	TES	UVM	Fold increase (log)	TES	UVM
Papillomaviridae	Betapapillomavirus 2	24	0	1,38	15,47	0
Picornaviridae	Aichivirus A	23372	33	2,85	89,86	22,05
	Enterovirus A	725	0	2,82	61,12	0
	Enterovirus B	902	0	2,96	44,00	0
	Enterovirus C	2098	0	3,32	62,10	0
Adenoviridae	Human mastadenovirus A	343	0	2,54	14,22	0
	Human mastadenovirus F	1333	0	3,12	34,66	0
Polomaviridae	BK polyomavirus	829	0	2,92	79,89	0
	Human polyomavirus 6	79	0	1,90	10,88	0
	JC polyomavirus	714	237	0,48	93,82	77,37
	WU polyomavirus	17	0	1,23	10,96	0
Astroviridae	Mamastrovirus 1	83857	64	3,12	99,25	30,45
Caliciviridae	Sapporo virus	421	10	1,62	62,07	57,33
	Norwalk virus	1820	8	2,36	88,72	11,73
Hepeviridae	Orthohepevirus A	781	0	2,89	24,21	0
Reoviridae	Rotavirus A	423	0	2,63	38,64	0

Fold increase is presented in logs of increase and breadth genome coverage in percentages.

Figure 1
Click here to download Figure: FIGURE 1 R1.pdf

Figure 1. Host distribution of viral species within families obtained by Target Enrichment Sequencing (TES) and Untargeted Viral Metagenomics (UVM) in WWTP A.



Colours indicate the amount of different viral species within each family, ranging from red (higher number) to green (lower number).

Figure 2
Click here to download Figure: FIGURE 2 R1.pdf

Figure 2. Diversity of specific viral groups (human adenoviruses, papillomaviruses and enteroviruses) obtained by Amplicon Deep Sequencing in the composite raw sewage sample A.

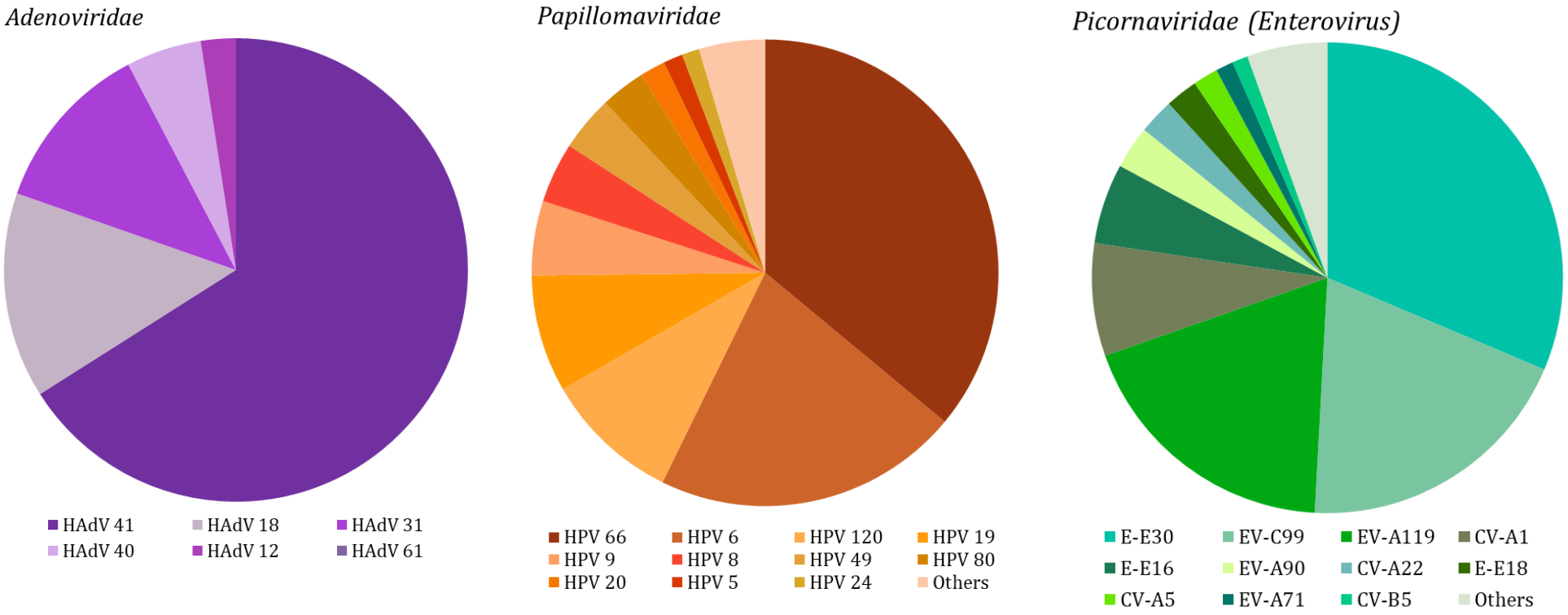


Figure 3

Click here to download Figure: FIGURE 3 R1.pdf

Figure 3. Diversity of *Adenoviridae* (a), *Papillomaviridae* (b) and *Picornaviridae* (c) families in the seasonal pooled raw sewage samples from WWTP B obtained by Amplicon Deep Sequencing. Data presented as percentage of viral species and number of reads.

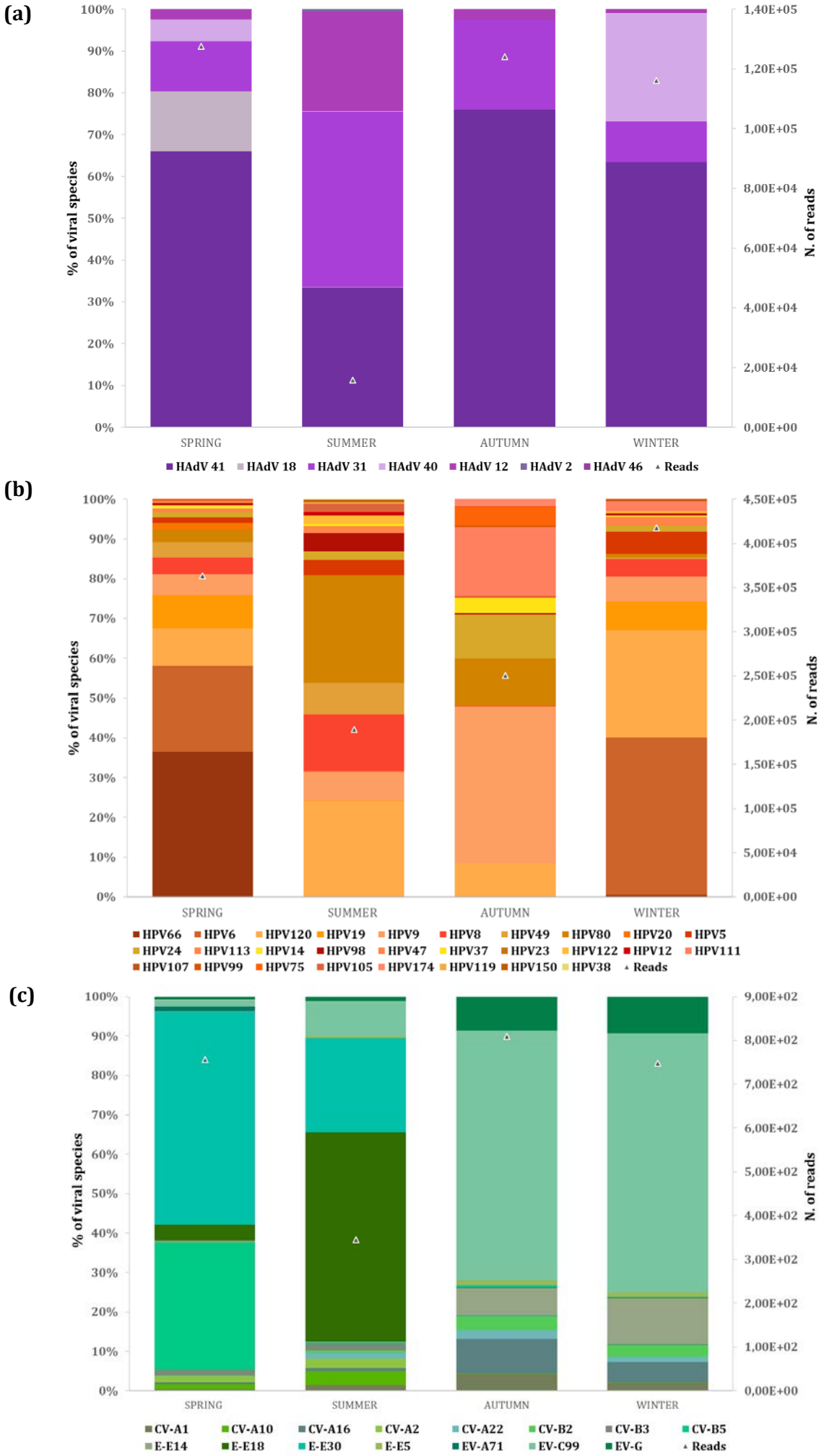


Figure 4
Click here to download Figure: FIGURE 4 R1.pdf

Figure 4. Phylogenetic tree based on enterovirus VP1 nucleotide sequences of EV-A71 C1 strains identified by ADS. Represented in bold are the sequences obtained in both WWTPs (BCN_A and BCN_B). The rest of the strains were the onex obtained from clinical cases by Sanger sequencing from Hospital Vall d'Hebrón (VH) and the ones related to the Germany outbreak in 2015 (those with KU prefix; Böttcher, 2016). Scale bar indicate nucleotide substitution per site. CV-A16 (KT327162) has been used as outgroup, with a distance of 0.511 to BCN A_08.