

Human Adenovirus Diversity in Water Samples Using a Next-Generation Amplicon Sequencing Approach

Leslie Ogorzaly¹ · Cécile Walczak¹ · MéliSSa Galloux¹ · Stéphanie Etienne² · Benoît Gassilloud² · Henry-Michel Cauchie¹

Received: 18 December 2014 / Accepted: 30 March 2015 / Published online: 28 April 2015
© Springer Science+Business Media New York 2015

Abstract This study aims to establish a straightforward and original workflow for high-throughput typing of human adenoviruses (HAdVs) in environmental samples. Occurrence of HAdVs in water is well documented worldwide, but data on diversity of HAdV types circulating in water are scarcely available. Here, the characterisation of viral particles was performed by determination of amplicon sequences using a next-generation sequencing (NGS) approach. Adenoviral DNA was either directly isolated from wastewater or river water concentrates or after a cell culture passage. Genome amplification targeted a hyper variable region of the hexon gene, allowing the discrimination of the 54 human adenoviral types described until now. After read generation on the benchtop MiSeq platform (Illumina), data were analysed using the Mothur software for identification of all HAdV species and types simultaneously present in a unique sample. NGS results showed a relatively wide HAdV diversity of up to six types in one sample, whereas Sanger sequencing always only retrieved the dominant one. Detected types included HAdV-1, HAdV-2, HAdV-3, HAdV-6, HAdV-12, HAdV-31, HAdV-40 and HAdV-41, HAdV-41 being the most abundant in tested samples. In addition, the influence of the cell line (A549 vs 293A cells) on the infectious HAdV typing results was clearly determined. The 293A appeared to be the

most suitable cell line allowing the detection of a larger diversity of infectious HAdVs and reflecting a more realistic initial species distribution than using the A549 cells. These findings demonstrated the feasibility of amplicon sequencing NGS approach to identify viruses in complex environmental water samples.

Keywords Next-generation sequencing · Human adenoviruses · Diversity · Typing · Wastewater · River water

Introduction

Human adenoviruses (HAdVs) are considered as one of the most abundant enteric virus groups in water. Numerous studies are available on their occurrence and their concentration in all types of water matrices all over the world (see review articles by Jiang 2006; Mena and Gerba 2009; Bofill-Mas et al. 2013). The detection of HAdVs in water is commonly carried out by PCR or real-time PCR using either consensus primers which target all HAdV types or specific primers for identifying exclusively the enteric types (HAdV-40 and HAdV-41). But paradoxically, the overall diversity of the HAdV types in environmental samples is still insufficiently documented (Lee and Kim-Sang 2010; Bibby and Peccia 2013b). Better knowledge of their occurrence in water could improve infectious risk assessment for humans exposed to specific HAdV types. Currently, 54 types of HAdVs are approved in the last report of the International Committee on Taxonomy of Viruses (Harrach et al. 2012). They are divided into seven HAdV species named HAdV-A to HAdV-G, which belong to the *Mastadenovirus* genus of the *Adenoviridae* family. Some other potential new types have been suggested

✉ Leslie Ogorzaly
leslie.ogorzaly@list.lu

¹ Department of Environmental Research and Innovation (ERIN), Luxembourg Institute of Science and Technology (LIST), 41, rue du Brill, 4422 Belvaux, Luxembourg

² French Agency for Food, Environmental and Occupational Health and Safety (ANSES), Nancy Laboratory for Hydrology, Water Microbiology Unit, 40, rue Lionnois, 54000 Nancy, France

(Walsh et al. 2011; Robinson et al. 2011; Liu et al. 2011; Matsushima et al. 2012) but are not yet accepted. Until now, sequencing efforts have resulted in the identification of only a small proportion of HAdV types in water samples, despite the fact that all of them can be shed in stool of infected people and then be discharged in wastewater treatment plants or directly into surface waters (Mena and Gerba 2009). The main reported water-associated types belong to the HAdV-A (HAdV-12, HAdV-31), HAdV-C (HAdV-1, HAdV-2, HAdV-5, HAdV-6) and the HAdV-F (HAdV-40, HAdV-41) species, with HAdV-41 and HAdV-2 being the most commonly detected types (Van Heerden et al. 2005; Muscillo et al. 2008; Kuo et al. 2009; Fong et al. 2010; Bofill-Mas et al. 2010; Wyn-Jones et al. 2011; Ogorzaly et al. 2013). Some types from HAdV-D (HAdV-8, HAdV-19) and HAdV-B (HAdV-3) have also been observed, but with very low frequency (Fong et al. 2010; Kokkinos et al. 2010; Bofill-Mas et al. 2010; Wyn-Jones et al. 2011). These data were mainly obtained by the direct sequencing (Sanger sequencing) of PCR products after a purification step. Cloning sequencing (insertion of PCR products into a plasmid vector prior to sequencing) was also used, but sparsely. The major drawback of these conventional sequencing approaches is that only a limited number of representative and especially the predominant sequences can be obtained, while water samples may contain multiple viral strains. Therefore, these methods are limited in output and inadequate for retrieving whole HAdVs diversity in environmental samples.

Today, the high-throughput next-generation sequencing (NGS) technology is an attractive and practical means for studying a viral pathogen's diversity in clinical samples as well as in environmental samples. Two NGS approaches can be applied in this framework: the random metagenomic which provides information on the full virome of a given sample or a gene-targeted sequencing which specifically detects viruses of interest. Advantages, relevance, drawbacks and limitations of both approaches were discussed in a comprehensive review (Wong et al. 2012). Recently published metagenomic data have confirmed the presence and the prevalence of HAdVs in wastewater, sewage sludge and biosolid viromes (Bibby et al. 2011; Bibby and Peccia 2013a; Aw et al. 2014). Using such an approach, information on family, genus and sometimes species of viruses can be gathered. However, more accurate data on types remains for the moment, difficult to obtain because of an insufficient sequencing depth. This limitation is probably a consequence of the presence of high concentrations of untargeted nucleic acids (plant viruses, bacteriophages or bacteria) conjointly sequenced with the target of interest (Hall et al. 2014). For the same reason, some discrepancies have been described between results from direct viral pathogen detection using real-time PCR and metagenomic

analysis (Bibby 2013; Bibby and Peccia 2013a). In order to increase the chance of obtaining detailed data on one or several viral pathogens of interest and an estimation of their proportion in the same sample, an amplicon sequencing approach appears to be preferable.

The objectives of this study were (i) the development of a next-generation amplicon sequencing assay for HAdVs and (ii) the detection and identification of HAdV diversity in wastewater and river water matrices. Viral DNA amplification was performed using a degenerated primer pair targeting the non-conserved hexon gene, which allows the discrimination of all 54 human types described until now. An Illumina benchtop MiSeq system was used for its ability to sequence long reads of 2×300 base pairs, which seems highly compatible for the suggested amplicon sequencing approach. In addition, due to its relevance and pertinence for public health issues and microbial risk assessment, typing of infectious particles was performed subsequently to a cell culture isolation and multiplication step. A broad variety of cell lines have been proposed, but the efficiency of viral replication on cell cultures varies with HAdV types. So, the impact of the choice of the cell line (A549 vs 293A) on the overall HAdV diversity observed in wastewater treatment plants is also evaluated and discussed.

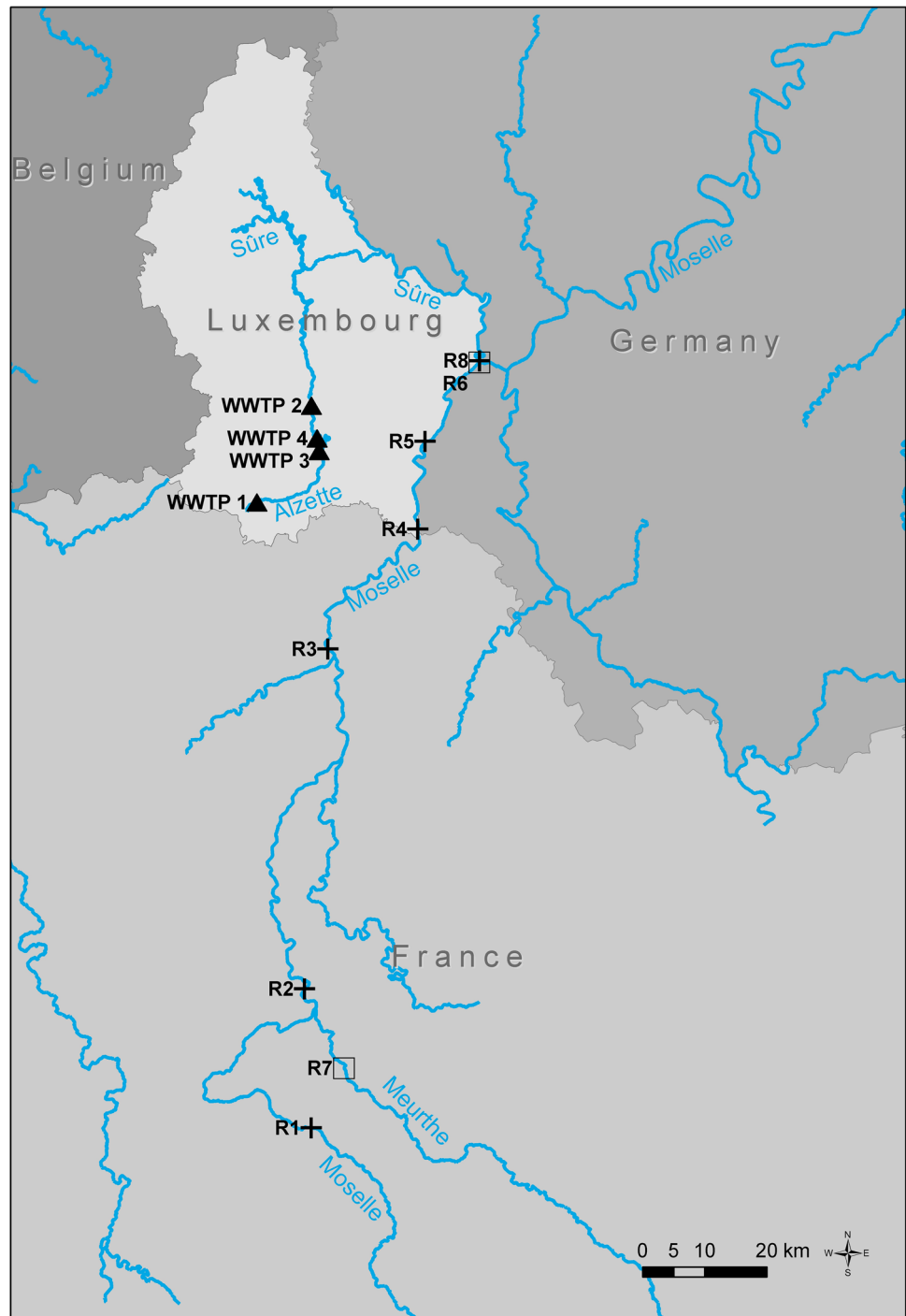
Materials and Methods

Water Samples

Influent and the effluent of four Luxembourgish wastewater treatment plants (WWTPs) located in Schifflange, Beggen, Hespérange and Bonnevoie (capacities of 90,000; 210,000; 26,000 and 60,000 inhabitant equivalents, respectively) were collected in March (Schifflange), April (Beggen and Hespérange) and May (Bonnevoie) 2013 (Fig. 1). Two hundred and four hundred millilitres of water samples from inlet and outlet of the treatment plants, respectively, were concentrated by ultracentrifugation as previously described (Skraber et al. 2011). The pellet was resuspended in 10 mL of Dulbecco's modified Eagle medium (D-MEM).

In addition, 24 samples of surface water were collected in the French and Luxembourgish parts of the Moselle River watershed (Fig. 1). Six sampling sites were dispatched along a 150-km-long section of the Moselle River and two additional sampling points were added on the two major tributaries, i.e. the Meurthe River and the Sûre River. Three successive sampling campaigns were performed in March, April and May 2014. Viral particles were concentrated from 20 L of acidified water (pH 3.5) using a glass wool method followed by an organic flocculation, as previously described (Wyn-Jones et al. 2011). The pellet was resuspended in 10 mL of D-MEM.

Fig. 1 Locations of the four wastewater treatment plants (filled triangle) and the sampling sites on the Moselle River (plus symbol R1–R6), the Meurthe River (open square R7) and the Sûre River (open square R8). WWTP 1 Schiffflange, WWTP 2 Beggen, WWTP 3 Hespérange, WWTP 4 Bonnevoie



A flowchart of analyses performed on water concentrates for HAdV detection and characterisation is presented in Fig. 2.

Cell Line Infection

The human lung carcinoma cell line A549 and the human embryonic kidney cell line 293A, commonly

adopted for HAdV detection, were used in this study. Both cell lines were used for wastewater samples (Fig. 2). Infectivity assays were performed according to both previously published ICC–quantitative real-time PCR (qPCR) protocols (Wyn-Jones et al. 2011; Ogorzaly et al. 2013). Briefly, cells were inoculated in a 25-cm² tissue culture flask. After a 4-day growing period, cell monolayers were infected with 1 mL of viral

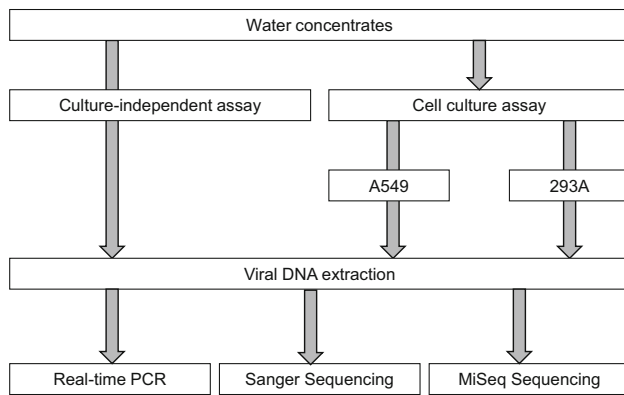


Fig. 2 Experimental flows to evaluate concentration, infective status and diversity of human adenovirus in water samples using culture-dependent and culture-independent methods. Culture-dependent assays were performed using both cell lines (A549 and 293A) for wastewater samples, while only 293A cells were used for river water samples

concentrates and incubated for 60 min at 37 °C for the attachment of the infectious viral particles on cells. A negative process control (ISO 22174:2005) was included in all our experiments, consisting in the inoculation of a cell monolayer with 1 mL of D-MEM medium. Afterwards, the viral inoculum was discarded, and cells were submitted to three successive washings to remove unbound viral particles. Finally, cells were inoculated with maintenance medium containing 1 % final concentration of an antibiotic–antimycotic solution (6,000 µg/mL of penicillin, 10,000 µg/mL of streptomycin and 25 µg/mL of amphotericin B, Life Technologies). A549 and 293A cells were incubated for 5 and 3 days, respectively, prior to viral DNA isolation. The presence of infectious HAdVs in river samples was evaluated using only the 293A cells, and the ICC–qPCR protocol was adapted for titre microplate format. Conditions and media used remained similar as described in the protocol of Ogorzaly et al. (2013).

DNA Isolation

All viral DNA isolations were carried out using the QIAamp Viral RNA mini kit (Qiagen) to a final volume of 100 µL of elution buffer. The initial sample quantity used for DNA extraction was adapted in accordance with the preceding sample preparation protocol. After cell culture, viral DNA was purified either from 140 µL of A549 cell supernatant after three freeze–thaw cycles or directly from 293A cells. For culture-independent detection, viral DNA was directly extracted from 140 µL and 1 mL of wastewater and surface water concentrates, respectively, using the QIAamp Viral RNA mini kit (Qiagen). A negative extraction control was added in each experiment.

Quantitative Real-Time PCR

Quantification of HAdV genome in samples was performed using consensus qPCR assays, using primers, probes and conditions of amplification described by Hernroth et al. (2002) and Jothikumar et al. (2005) for wastewater and river water samples, respectively. The qPCR standard curves were generated using plasmids DNA having the cloned target regions. The validity of the qPCR results was confirmed using an appropriate set of controls (positive and negative qPCR controls) at each cyclor run. Also, qPCR inhibition was evaluated either by the addition of a known concentration target plasmid DNA or by the dilution (10- and 100-fold) of DNA extracts prior to amplification.

Sanger Sequencing

In order to identify the major HAdV type occurring in each water sample, DNA extracts from all samples were amplified and analysed by Sanger sequencing according to a previously described protocol (Ogorzaly et al. 2013).

NGS Amplicon Production and Sequencing

The first step used PCR to amplify HAdV DNA template out of DNA sample using a region of interest-specific primers with attached overhang adapters. This amplification was performed using the degenerated hex1deg–hex2deg primer pair (Allard et al. 2001), which targets a hyper variable region of the hexon gene allowing the discrimination of the 54 HAdV types described until now. A 50-µL PCR mix was prepared for each sample and contained 5 µL of 10× High Fidelity PCR buffer without magnesium salts (Platinum *Taq* DNA Polymerase High Fidelity kit, Invitrogen), 2 µL of 50 mM MgSO₄ (Platinum *Taq* DNA Polymerase High Fidelity kit, Invitrogen), 2.5 µL of 2.5 mM dNTPs each (Invitrogen), 0.2 µL of Platinum *Taq* DNA polymerase High Fidelity (Platinum *Taq* DNA Polymerase High Fidelity kit, Invitrogen), 0.5 µL of each 50 mM modified primer, 29.3 µL of PCR grade water and 10 µL of DNA suspension. Thermal profile was 94 °C for 60 s, followed by 25 cycles at 94 °C for 30 s, 55 °C for 30 s and 68 °C for 60 s. After amplification, PCR products were purified using AMPure XP beads (Beckman Coulter) according to the manufacturer’s protocol with a ratio of 1.8:1 beads to DNA. The concentration of purified amplicons was measured using the Qubit fluorometer (Invitrogen) and the Qubit dsDNA HS assay kit (Invitrogen).

The second step used PCR to fix, at each extremity of the purified amplicons, dual indices and Illumina sequencing primers to obtain barcoded amplicons ready for sequencing. Fixation was carried out using the overhang

adapters previously attached to amplicons. The unique index combination enables the subsequent identification of each sample. Amplification was performed in a 50 μL reaction mixture containing 5 μL of PCR products (1 $\text{ng}/\mu\text{L}$), 25 μL of Q5 Hot Start High Fidelity 2 \times Master Mix (Biolabs), 5 μL of Nextera XT Index 1 (N7–, Illumina), 5 μL of Nextera XT Index 2 (S5–, Illumina) and 10 μL of PCR grade water. Thermal profile was 95 $^{\circ}\text{C}$ for 30 s, followed by 12 cycles at 95 $^{\circ}\text{C}$ for 10 s, 55 $^{\circ}\text{C}$ for 30 s and 72 $^{\circ}\text{C}$ for 30 s, and a finishing step at 72 $^{\circ}\text{C}$ for 5 min. After amplification, barcoded amplicons were cleaned using AMPure XP beads (Beckman Coulter) as described above. The length of purified barcoded amplicons was assessed by capillary electrophoresis (Agilent 2100 Bioanalyser, Agilent Technologies) from 1 μL of the PCR products using the Agilent DNA 1000 kit (Agilent Technologies). Also, purified barcoded amplicons were diluted 1:4000 or 1:10,000 for quantification by qPCR using the KAPA Library Quantification kit (KAPA Biosystems) with dsDNA standard concentrations in a range from 2×10^{-4} to 20 pM, according to the manufacturer's instructions.

All quantified amplicons were mixed in equimolar amounts to obtain a final DNA library concentration of 4 nM. The final library concentration was controlled by qPCR using the KAPA Library Quantification Illumina kit. The DNA library was then denatured and diluted as recommended by Illumina (MiSeq Reagent kit v3-reagent preparation guide) and loaded at 15 pM on a MiSeq flow cell with 5 % PhiX spiked in. Paired-end sequencing was performed with 301 cycles using the Generate FASTQ workflow. After sequencing, MiSeq reporter software automatically trimmed Illumina adapters and demultiplexed data to provide two FASTQ files (read 1 and read 2) per sample.

Bioinformatics for HAdV-Type Identification

The sequences dataset was analysed using the program Mothur (Schloss et al. 2009) following the standard operating procedure (SOP) for MiSeq provided on the Mothur platform (http://www.mothur.org/wiki/MiSeq_SOP) (Kozich et al. 2013). Briefly, after contig generation, sequences were removed from consideration if they did not have a length ranging between 290 and 320 base pairs. Unique sequences from all samples were merged and aligned to a reference alignment. The reference alignment file was created using the official sequence for each approved HAdV type and CLC Main Workbench 6 software (CLC Bio). The aligned sequences were then treated to trim the PCR primers and to discard sequencing mistakes and chimaeras. Chimaeras were removed using the Uchime algorithm implemented in Mothur. All remaining sequences were finally classified. A value of 99 % of

similarity was fixed for HAdV type identification, except for the distinction of types belonging to the HAdV-C species. In this case, 70 % of similarity was used as an identification criterion.

Results

Abundance and Diversity of HAdV in Wastewater Samples

Eight wastewater samples (4 influents and 4 effluents) from four treatment plants were used in this study. Using a culture-independent assay, genetic material of HAdVs was detected in all tested wastewater samples. The DNA concentrations ranged from 6.3×10^4 to 1.1×10^6 and 3.2×10^4 to 2.8×10^5 genome copies/L for influents and effluents, respectively. Sanger sequencing of amplicons generated by direct PCR showed the presence of both HAdV-31 and HAdV-41 types, while next-generation sequencing results showed the presence of adenoviral DNA belonging to the HAdV-A (HAdV-31, HAdV-12), HAdV-C (HAdV-2) and HAdV-F (HAdV-40, HAdV-41) species (Fig. 3). The most abundant species in wastewater of the four treatment plants was the F, and especially the HAdV-41 with a relative abundance of more than 80 %.

All samples tested were also positive for the occurrence of infectious adenoviral particles. HAdVs isolated from cell culture using A549 and 293A cells were sequenced on the MiSeq platform in order to assess the effect of the cell line on identified HAdV diversity. Globally, the diversity picture obtained using both cell lines was remarkably different (Fig. 3). Of the A549 reads, the majority of contigs (superior to 95 % regardless of the sample site) were assigned to HAdV-C with a high significant value, and more particularly as HAdV-2 (Fig. 3). In samples from the WWTP 2 (Beggen), a small number of reads were annotated as HAdV-31 (2.3 %) belonging to the HAdV-A species, and as HAdV-3 (1.4 %) belonging to the HAdV-B species. The main divergence observed using the 293A cells was the broad detection of infectious HAdV-F specifically, permitted by the 293A cells. Both HAdV-41 and HAdV-40 were represented, with the HAdV-41 being the major type. Also, more species and more types were observed after a preliminary isolation and culture step on 293A cells. The three species HAdV-A, HAdV-C and HAdV-F were very well characterised using the 293A cells. In all samples together, six different types were observed, i.e. HAdV-1, HAdV-2, HAdV-12, HAdV-31, HAdV-40 and HAdV-41 (Fig. 3). They were also found simultaneously in water samples coming out of inlet and outlet of the WWTP 2.

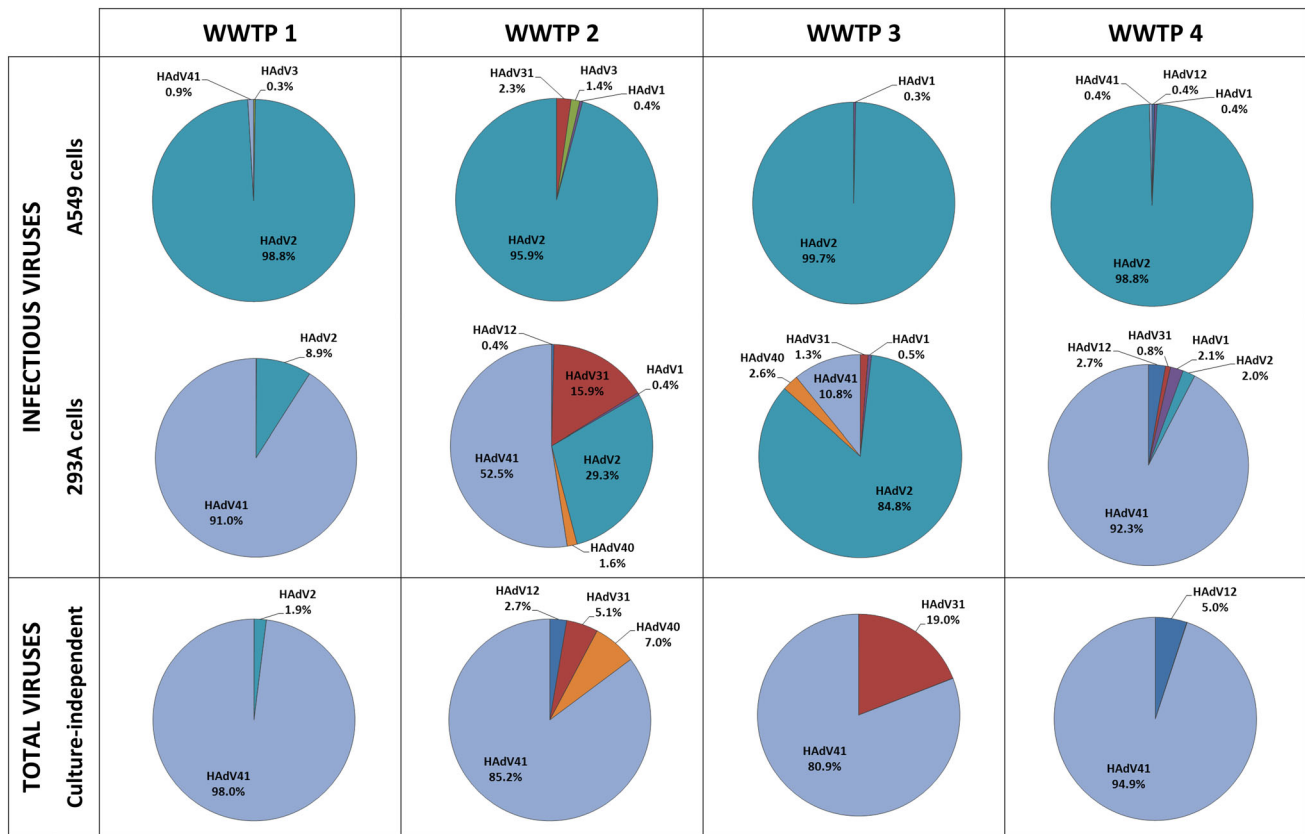


Fig. 3 Distribution of human adenovirus types identified in wastewater samples isolated directly from water (total viruses) and from A549 and 293A cells (infectious viruses). Data from inlet and outlet of each wastewater treatment plant were pooled

Interestingly, Sanger sequencing results were highly consistent with NGS results for all wastewater samples analysed regarding the predominant HAdV type identified. Indeed, the only HAdV type identified by Sanger sequencing was also at the same time the most abundant type observed by MiSeq sequencing.

Abundance and Diversity of HAdV in River Water Samples

Concerning river water samples, the genome of HAdV was detected in 87.5 % (*n* = 24) of the samples (Table 1). However, in most of the samples, HAdV quantities were closed to the limit of quantification (LOQ) of the qPCR assay, hindering a determination of the genome copy concentration. Among the 21 samples declared positive, seven samples were quantifiable (superior to LOQ estimated to 200 genome copies/L) and exhibited an average concentration of $2.9 \pm 0.5 \times 10^2$ genome copies/L. qPCR inhibition was evaluated by the addition of plasmid DNA having the target region at a known concentration in each sample, and it was estimated to be less than 30 % in all samples (data not shown). Due to the low viral DNA quantities in our samples, no amplicon was obtained by

qualitative PCR using the hex1deg-hex2deg primers set, precluding the MiSeq amplicon sequencing. To acquire a hexon gene amplicon, a nested-PCR was carried out using the nehex3deg-nehex4deg primer set. Only seven amplified DNA fragments were obtained from the 21 samples that were positive by consensus qPCR, including six HAdV-41 and one HAdV-3 (Sanger sequencing).

About half of the river samples (54.2 %, *n* = 24) were found positive for the occurrence of infectious particles using the 293A cells. The MiSeq amplicon sequencing workflow was applied on these positive samples. One hundred percent of obtained reads were annotated as HAdV-41 in 12 out of 13 samples. For the last one, two different types were determined together, the HAdV-6 and the HAdV-41. Identical results were found by applying Sanger sequencing because of a moderate quantity of HAdV and a poor diversity in these river water samples.

Discussion

Although today the metagenomic approach is preferred for determining the viral diversity in wastewater, the findings in this study establish that the amplicon sequencing

Table 1 River water results for the detection and characterisation of human adenoviruses by quantitative real-time PCR, nested-PCR followed by Sanger and/or MiSeq amplicon sequencing

River	Culture-independent detection				Culture-dependent detection			
	Positives/total number of tested samples	Genome (copies/L) ^a	MiSeq sequencing	Sanger sequencing hex1-hex2	Sanger sequencing nehex3-nehex4 (positives/total number of tested samples)	Positives/total number of tested samples	MiSeq sequencing (% of reads associated to HAdV types)	Sanger sequencing hex1-hex2 (positives/total number of tested samples)
Moselle	11/18	312 ± 50	na	na	HAdV-41 (4/15)	10/18	HAdV-41 (100)	HAdV-41 (10/10)
Meurthe	2/3	240	na	na	HAdV-41 (2/3)	2/3	HAdV-6 (59)	HAdV-6 (1/2)
Sûre	1/3	265 ± 50	na	na	HAdV-3 (1/3)	1/3	HAdV-41 (100)	HAdV-41 (1/1)
All	87.5 %					54.2 %		

LOQ limit of quantification, na no amplicon

^a Average and standard deviation of genome copies detected by real-time PCR

approach can also be highly suitable and appealing for the rapid detection and identification of viruses of interest in complex water matrices. Using such an approach, the detection specificity is favoured over the production of high data quantity for obtaining a maximum of relevant information from the millions of reads produced in the sequencing run (Lange et al. 2014). To the best of our knowledge, this is the first work which provides information on HAdV-type occurrence and prevalence in wastewaters and surface waters using a next-generation amplicon sequencing assay. An analogous study has been reported using another NGS platform, but the results were restricted to HAdV species identification (Bibby and Pectia 2013b). The data treatment developed in the present study allows to assign all reads generated by NGS to a single HAdV species and type with a very high significant value (99 % of similarity selected as critical value). One difficulty was, however, encountered for the distinction of types belonging to the HAdV-C species. Reads assigned to HAdV-2, HAdV-6 and HAdV-1 strains presented a similarity of approx. 70, 80 and 99 %, with the reference-type strain (J01917, HC492785 and AF534906, Genbank accession number), respectively. But all of these amplicons were unequivocally assigned to HAdV-C species using the critical value of 99 % of similarity. Other authors also mentioned such an observation in their study (Lee et al. 2004). One possible explanation might be the high genomic similarity of the four human types from the HAdV-C species. Also, reads obtained in this study were only examined in contrast to the reference sequence, but many other sequences are available in databases for one given strain.

Using the suggested MiSeq workflow, the distribution and the relative abundance of HAdV types observed in the analysed samples are in agreement with those retrieved from scientific articles taken all together. Overall, only some HAdV types from the 54 types currently described are detected in our wastewater and freshwater samples. The HAdV types 2, 12, 31, 40 and 41 being the most commonly represented, with some occurrences of HAdV-1, HAdV-3 and HAdV-6. Also, the high prevalence of the HAdV-41 highlighted here, especially in river water, is in accordance with what has been previously reported (Xagorarakis et al. 2007; Haramoto et al. 2007). Thus, the workflow used here is perfectly suited for evaluating HAdV diversity in wastewater samples and for typing infectious particles. However, some limitations have been observed for river samples sparsely contaminated by viruses. No hex1deg-hex2deg amplicon was obtained, despite the genomic signal in real-time PCR and detection by culture assay. The low viral DNA quantities present in the concentrated water samples and the lack of sensibility of the qualitative PCR assay can be considered as responsible of the failure of

amplicon generation using hex1deg-hex2deg primer pair. Consequently, a nested-PCR had to be applied to obtain a PCR fragment usable for Sanger sequencing. This latter approach allowed obtain information for some additional samples but unfortunately not for all of them. In this particular context, a preliminary enrichment step can be an interesting alternative (Hall et al. 2014), as demonstrated here. The use of a cell culture step prior to the HAdV detection and characterisation by NGS has a double asset. On the one hand, it provides information on infectious particles only, which can be directly related to a public health risk assessment. On the other hand, multiplication inside host cells allows drastically increase the viral DNA quantity for subsequent typing analysis. However, this preliminary culture step may potentially alter the original species/types distribution according to the experimental procedure applied, especially regarding the cell line type or the culture duration.

Currently, the detection of infectious HAdVs in environmental samples is widely performed using either the A549 or the HEK 293A cells. As expected, the results reported reveal completely disparate HAdV diversity profiles according to the cell line used, that confirms and reinforces previous studies (Jiang et al. 2009; Ogorzalay et al. 2013). Our results clearly demonstrate that the A549 cells promote the multiplication of HAdV types belonging to the HAdV-C species at the expense of others. The presence of some annotated sequences from HAdV-31 and HAdV-3 types suggests that HAdV-A and HAdV-B species can also be detected using the A549 cells, albeit to a lesser extent. Data from the literature described that A549 supported the replication of the HAdV-C, HAdV-B, HAdV-D species but not the replication of HAdV-F (Jiang et al. 2009). Based on our results, no conclusion can be done on the HAdV-D species since it was not detected during the present study. In accordance with our results, the 293A cell line enables the multiplication of the HAdV-A, HAdV-C and HAdV-F species in the same way. Other work also reported the replication of the HAdV-D (Jiang et al. 2009) on this cell line. In addition, the relative abundance of each HAdV type observed using the 293A cells is closely related to the diversity observed by culture-independent typing; however, the results are not directly comparable (only infectious particles are identified in the first way, whereas DNA fragments from infectious and non-infectious viruses are conjointly characterised in the other way). A high proportion of infectious particles in our samples, as established using the infectivity assay, is a quite probable explanation of this likeness. The 293A appears to be the most suitable cell line to allow the detection of a wide diversity of HAdVs occurring in water samples, giving thus a more realistic types/species distribution than using the A549 cell line. A recently published study described the use of a

metagenomic approach on the MiSeq Illumina platform to identify enteric pathogenic viruses isolated from wastewater with or without a cell culture step on A549 cells (Aw et al. 2014). In the latter study, no contig was identified as HAdV-F after a culture step on A549 cells, while in the original sample, about 15 % of the annotated contigs were associated to this enteric species, using a direct assay. Unfortunately, no information was provided on the infective state of the HAdV-F, preventing conclusive interpretation. However, it is important to note that this observation is largely in agreement with the findings presented in the present work, and supports our conclusion on the inability of A549 cells to give a clear overview of the HAdV diversity in a water sample.

Results obtained in this study are very consistent with previous data on HAdV types circulating in water matrices, confirming the high prevalence of enteric types. The accuracy of sequences generated by this NGS method was demonstrated by comparing sequences to those generated by Sanger PCR amplicon sequencing. The HAdV type identified by Sanger sequencing and the most abundant type observed by NGS were always exactly the same in all samples, regardless the water type and the sample preparation protocol. Moreover, alignment of the Sanger and MiSeq generated sequences revealed very high level of matching. The major asset of MiSeq sequencing over the Sanger approach is the ability to simultaneously retrieve multiple sequences belonging to different HAdV types from a unique sample. The overall diversity of HAdV in aquatic environment is thus accessible, in particular the less represented types. The described NGS workflow appears to be an attractive alternative because it is easier to set up, more informative and more cost-efficient, as a high number of samples can be analysed simultaneously.

In conclusion, a significant output of our study is the expansion of the available information on the HAdV-type identification and diversity in complex environmental water thanks to the next-generation amplicon sequencing approach. This use for water sample analysis has the potential to provide relevant, accurate and specific information about the diversity of pathogens of concern with a unique sequencing protocol. The sample preparation step remains however a critical factor for the effective identification of viruses, as viral nucleic acids need to be present in high enough concentration to guarantee the detection. Similar studies could be conducted to monitor simultaneously different enteric viruses in order to improve the knowledge about the pathogenic viruses circulating in aquatic environments.

Acknowledgments The present work was carried out in the framework of the LorLux network (French name, RésEAU LorLux) supported by the National Research Fund, Luxembourg, and the

Lorraine region, France. The authors would like to thank D. Collard and A. Tavares Furtado for their excellent technical assistance and Dr. M. Calusinska for her precious help in bioinformatics data treatment. Dr. C. Penny and Dr. L. Hoffmann are also acknowledged for final reading of the manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Standards The manuscript does not contain clinical studies or patient data.

Informed Consent The manuscript does not contain clinical studies or patient data.

References

- Allard, A., Albinsson, B., & Wadell, G. (2001). Rapid typing of human adenoviruses by a general PCR combined with restriction endonuclease analysis. *Journal of Clinical Microbiology*, *39*(2), 498–505.
- Aw, T. G., Howe, A., & Rose, J. B. (2014). Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *Journal of Virological Methods*, *210*, 15–21.
- Bibby, K. (2013). Metagenomic identification of viral pathogens. *Trends in Biotechnology*, *31*(5), 275–279.
- Bibby, K., & Peccia, J. (2013a). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental Science and Technology*, *47*(4), 1945–1951.
- Bibby, K., & Peccia, J. (2013b). Prevalence of respiratory adenovirus species B and C in sewage sludge. *Environmental Sciences: Processes and Impacts*, *15*(2), 336–338.
- Bibby, K., Viau, E., & Peccia, J. (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Letters in Applied Microbiology*, *52*(4), 386–392.
- Bofill-Mas, S. Á., Calgua, B., Clemente-Casares, P., La Rosa, G., Iaconelli, M., Muscillo, M., et al. (2010). Quantification of human adenoviruses in European recreational waters. *Food and Environmental Virology*, *2*(2), 101–109.
- Bofill-Mas, S., Rusñol, M., Fernandez-Cassi, X., Carratalà, A., Hundesa, A., & Girones, R. (2013). Quantification of human and animal viruses to differentiate the origin of the fecal contamination present in environmental samples. *BioMed Research International*, *2013*, 192089.
- Fong, T. T., Phanikumar, M. S., Xagorarakis, I., & Rose, J. B. (2010). Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a Michigan River. *Applied and Environmental Microbiology*, *76*(3), 715–723.
- Hall, R. J., Wang, J., Todd, A. K., Bissielo, A. B., Yen, S., Strydom, H., et al. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*, *195*, 194–204.
- Haramoto, E., Katayama, H., Oguma, K., & Ohgaki, S. (2007). Quantitative analysis of human enteric adenoviruses in aquatic environments. *Journal of Applied Microbiology*, *103*(6), 2153–2159.
- Harrach, B., Benkö, M., Both, G., Brown, M., Davison, A., Echavarría, M., et al. (2012). Family—Adenoviridae. In A. King, M. Adams, E. Carstens, & E. Lefkowitz (Eds.), *Virus taxonomy: Ninth report of the international committee on taxonomy of viruses* (pp. 125–141). San Diego: Elsevier.
- Hernroth, B. E., Conden-Hansson, A. C., Rehnstam-Holm, A. S., Girones, R., & Allard, A. K. (2002). Environmental factors influencing human viral pathogens and their potential indicator organisms in the blue mussel, *Mytilus edulis*: The first Scandinavian report. *Applied and Environmental Microbiology*, *68*(9), 4523–4533.
- Jiang, S. C. (2006). Human adenoviruses in water: Occurrence and health implications: A critical review. *Environmental Science and Technology*, *40*(23), 7132–7140.
- Jiang, S. C., Han, J., He, J. W., & Chu, W. (2009). Evaluation of four cell lines for assay of infectious adenoviruses in water samples. *Journal of Water and Health*, *7*(4), 650–656.
- Jothikumar, N., Cromeans, T. L., Hill, V. R., Lu, X., Sobsey, M. D., & Erdman, D. D. (2005). Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Applied and Environmental Microbiology*, *71*(6), 3131–3136.
- Kokkinos, P., Filippidou, S., Karlou, K., & Vantarakis, A. (2010). Molecular typing of enteroviruses, adenoviruses, and Hepatitis A viruses in untreated and treated sewage of a biological treatment plant in Greece. *Food and Environmental Virology*, *2*(2), 89–96.
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120.
- Kuo, D. H., Simmons, F., & Xagorarakis, I. (2009). A new set of PCR assays for the identification of multiple human adenovirus species in environmental samples. *Journal of Applied Microbiology*, *107*(4), 1219–1229.
- Lange, V., Böhme, I., Hofmann, J., Lang, K., Sauter, J., Schöne, B., et al. (2014). Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*, *15*(1), 63.
- Lee, C., & Kim-Sang, J. (2010). Single-strand conformation polymorphism analysis for the study of adenoviral diversity in urban rivers. *Journal of Virological Methods*, *165*(2), 145–150.
- Lee, C., Lee, S. H., Han, E., & Kim, S. J. (2004). Use of cell culture-PCR assay based on combination of A549 and BGMK cell lines and molecular identification as a tool to monitor infectious adenoviruses and enteroviruses in river water. *Applied and Environmental Microbiology*, *70*(11), 6695–6705.
- Liu, E. B., Ferreyra, L., Fischer, S. L., Pavan, J. V., Nates, S. V., Hudson, N. R., et al. (2011). Genetic analysis of a novel human adenovirus with a serologically unique hexon and a recombinant fiber gene. *PLoS ONE*, *6*(9), e24491.
- Matsushima, Y., Shimizu, H., Kano, A., Nakajima, E., Ishimaru, Y., Dey, S. K., et al. (2012). Novel human adenovirus strain, Bangladesh. *Emerging Infectious Diseases*, *18*(5), 846–848.
- Mena, K. D., & Gerba, C. P. (2009). Waterborne adenovirus. *Reviews of Environmental Contamination and Toxicology*, *198*, 133–167.
- Muscillo, M., Pourshaban, M., Iaconelli, M., Fontana, S., Di Grazia, A., Manzara, S., et al. (2008). Detection and quantification of human adenoviruses in surface waters by nested PCR, TaqMan real-time PCR and cell culture assays. *Water, Air, and Soil pollution*, *191*(1), 83–93.
- Ogorzal, L., Cauchie, H. M., Penny, C., Perrin, A., Gantzer, C., & Bertrand, I. (2013). Two-day detection of infectious enteric and non-enteric adenoviruses by improved ICC-qPCR. *Applied Microbiology and Biotechnology*, *97*(9), 4159–4166.
- Robinson, C. M., Singh, G., Henquell, C., Walsh, M. P., Peigue-Lafeuille, H., Seto, D., et al. (2011). Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. *Virology*, *409*(2), 141–147.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541.
- Skraber, S., Langlet, J., Kremer, J. R., Mossong, J., De Landtsheer, S., Even, J., et al. (2011). Concentration and diversity of noroviruses

- detected in Luxembourg wastewaters in 2008-2009. *Applied and Environmental Microbiology*, 77(15), 5566–5568.
- Van Heerden, J., Ehlers, M. M., Heim, A., & Grabow, W. O. K. (2005). Prevalence, quantification and typing of adenoviruses detected in river and treated drinking water in South Africa. *Journal of Applied Microbiology*, 99(2), 234–242.
- Walsh, M. P., Seto, J., Liu, E. B., Dehghan, S., Hudson, N. R., Lukashev, A. N., et al. (2011). Computational analysis of two species C human adenoviruses provides evidence of a novel virus. *Journal of Clinical Microbiology*, 49(10), 3482–3490.
- Wong, K., Fong, T. T., Bibby, K., & Molina, M. (2012). Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environment International*, 45, 151–164.
- Wyn-Jones, A. P., Carducci, A., Cook, N., D'Agostino, M., Divizia, M., Fleischer, J., et al. (2011). Surveillance of adenoviruses and noroviruses in European recreational waters. *Water Research*, 45(3), 1025–1038.
- Xagorarakis, I., Kuo, D. H. W., Wong, K., Wong, M., & Rose, J. B. (2007). Occurrence of human adenoviruses at two recreational beaches of the great lakes. *Applied and Environmental Microbiology*, 73(24), 7874–7881.