

SCIENTIFIC REPORTS



OPEN

Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters

Akihiko Hata^{1,2}, Masaaki Kitajima³, Eiji Haramoto⁴, Suntae Lee¹, Masaru Ihara¹, Charles P. Gerba⁵ & Hiroaki Tanaka¹

Human astroviruses are associated with gastroenteritis and known to contaminate water environments. Three different genetic clades of astroviruses are known to infect humans and each clade consists of diverse strains. This study aimed to determine the occurrence and genetic diversity of astrovirus strains in water samples in different geographical locations, i.e., influent and effluent wastewater samples (n = 24 each) in Arizona, U.S., and groundwater (n = 37) and river water (n = 14) samples collected in the Kathmandu Valley, Nepal, using next-generation amplicon sequencing. Astrovirus strains including rare types (types 6 and 7 classical human astroviruses), emerging type (type 5 VA-astroviruses), and putative recombinants were identified. Feline astrovirus strains were collaterally identified and recombination between human and feline astroviruses was suggested. Classical- and VA-astroviruses seemed to be prevalent during cooler months, while MLB-astroviruses were identified only during warmer months. This study demonstrated the effectiveness of next-generation amplicon sequencing for identification and characterization of genetically diverse astrovirus strains in environmental water.

Astroviruses (AstVs) are classified into the family *Astroviridae* and known to infect a wide range of mammalian and avian hosts including humans¹. AstVs possess icosahedral capsid with a diameter of 28–31 nm packaging single-stranded positive-sense RNA genome of 6.2–7.8 kb in length¹. The genome contains three open reading frames (ORFs), i.e., ORF1a, ORF1b, and ORF2. ORF1a and ORF1b encode nonstructural proteins such as serine protease and RNA dependent RNA polymerase, while ORF2 encodes structural proteins¹. To date, genetically diverse AstV strains have been identified. The strains so called human AstVs are regarded as “classical” AstVs (CAstVs), which are known as an important cause of infantile viral gastroenteritis. Their infections occur mainly sporadically and contribute to 2–10% of gastroenteritis cases^{2–4}. Recent investigations identified emerging human AstVs belonging to two clades (MLB- and VA-clades), which are genetically distinct from CAstVs. Genetically diverse strains have also been found within each clade. CAstVs, MLB-AstVs, and VA-AstVs are currently divided into 8, 3, and 5 genotypes, respectively^{1,5}. Human AstVs including the novel clades are regarded as enteric pathogens, but recent studies suggested that they also affect the central nervous system^{6–8}. Type 1 VA-AstVs are strongly suggested to be associated with encephalitis in immunocompromised patients⁷. Compared to other human enteric viruses such as noroviruses and rotaviruses, human AstVs are not well studied⁹. Therefore, especially for MLB- and VA-AstVs, their characteristics such as prevalence and epidemiology have not been fully elucidated.

¹Research Center for Environmental Quality Management, Kyoto University, Shiga, Japan. ²Department of Urban Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan. ³Division of Environmental Engineering, Faculty of Engineering, Hokkaido University, Hokkaido, Japan. ⁴Interdisciplinary Center for River Basin Environment, Graduate Faculty of Interdisciplinary Research, University of Yamanashi, Yamanashi, Japan. ⁵Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, Arizona, USA. Correspondence and requests for materials should be addressed to A.H. (email: hata@env.t.u-tokyo.ac.jp)

Clade	Genotype	U.S.		Nepal	
		Influent wastewater	Treated effluent wastewater	Groundwater	River Water
		(n = 24)	(n = 24)	(n = 37)	(n = 14)
CAstV ^a	Type 1	16	11	1	7
	Type 2	11	6	1	5
	Type 3	4	4	0	3
	Type 4/8	18	13	1	6
	Type 5	10	6	1	5
	Type 6	2	1	0	4
	Type 7	2	1	0	6
	FAstV ^b	5	6	1	0
Putative recombinant ^c	4	1	0	0	
	Any type (% positive)	18 (75%)	15 (63%)	2 (5.4%)	8 (57%)
MLB-AstV	Type 1	1	1	0	4
	Type 2	1	0	0	3
	Type 3	0	0	1	3
	Putative recombinant	0	0	0	2
		Any type (% positive)	1 (4.2%)	1 (4.2%)	1 (2.7%)
VA-AstV	Type 1	10	2	0	4
	Type 2	10	3	0	4
	Type 3	9	2	0	1
	Type 4	0	0	0	0
	Type 5	0	0	0	1
		Any type (% positive)	13 (54%)	5 (21%)	0 (0.0%)

Table 1. Positive rates of each genotype of human astroviruses from each sample matrix. ^aCAstV, classical human astrovirus. ^bFAstV, feline astrovirus. ^cNumber of OTU that were suggested to be recombinant by Simplot analysis.

Enteric viruses including human AstVs are excreted in the feces from infected individuals and contaminate the water environment directly or via wastewater treatment facilities¹⁰. Clinical surveillance is widely applied to understand the prevalence of enteric viruses^{2,3,5,11–13}. However, enteric virus infections occasionally result in asymptomatic cases that are not usually identified by clinical surveillance^{14,15}. Presence of enteric viruses in water environments indicates the presence of infected individuals including those asymptotically infected and potential infection risk via water. Thus, investigation of viruses in environmental water has significance in view of epidemiology and risk management. During these two decades since the first identification of human AstVs in the water environment in Spain¹⁶, only a few studies have focused on human AstVs in water¹. Particularly, presence of MLB- and VA-AstVs in environmental water samples has been reported by only two studies conducted in Japan and Uruguay^{17,18}. Both studies revealed that genetically diverse human AstV strains are present in the environment^{17,18}. This motivated us to investigate the presence of human AstVs in other regions.

Most of the previous studies analyzed AstV nucleotide sequences in environmental samples using Sanger sequencing, by which limited numbers of sequence reads can be obtained^{17–23}. Recently, next-generation sequencing technologies became available, which potentially allows investigation of viral genetic diversity in a given environmental sample with much higher throughput and greater sequencing depth. As previously reported, genetically diverse viral strains are co-present in fecally contaminated water^{17,21,24–26}. Next-generation amplicon sequencing can be an ideal technique to determine genetic diversity of a target virus in a water sample in depth.

This study aimed to investigate the prevalence and genetic diversity of human AstVs including the novel clades in water samples in the U.S. and Nepal. We also intended to demonstrate the usefulness of next-generation amplicon sequencing approach to reveal the presence of emerging/rare AstV strains in complex environmental samples.

Results

Occurrence of human AstVs in water samples. In this study, we sequenced 43, 7, and 23 of CAstV, MLB-, and VA-AstV positive nested-PCR amplicons, respectively, with three separate NGS runs (Table S1 in the Supplementary Information). A total of 79,957,568 reads were obtained. Among them, 10,367,900, 185,136, and 556,473 reads showed >70% similarity to CAstV, MLB-, and VA-AstV reference strains, respectively. Table 1 summarizes the occurrence of human AstVs belonging to CAstV, MLB-, and VA-clades in waste, ground, and river water samples. From wastewater samples collected in U.S. (n = 24 each for influent and treated effluent samples), CAstVs were identified with relatively high frequencies, 75% (18/24) and 63% (15/24) from influent and treated effluent samples, respectively, while MLB-AstVs were identified in only 4.2% (2/48) of the wastewater samples. VA-AstVs were identified in 54% (13/24) and 21% (5/24) of influent and treated effluent samples, respectively. WWTP-A and -B utilize different treatment processes, i.e., conventional activated sludge process and biological

tricking filter process, respectively. There was no clear difference in the detection frequencies of human AstVs in the effluent samples between these two WWTPs; CAstV, MLB-, and VA-clades were identified from 7, 0, and 2 ($n = 12$ each) effluent samples, respectively, at WWTP-A, and 8, 1, and 3 ($n = 12$ each) effluent samples, respectively, at WWTP-B.

Groundwater samples collected in Nepal ($n = 37$) resulted in low detection frequencies for human AstVs, only 2 (5.4%) and 1 (2.7%) samples were positive for CAstV and MLB-AstV, respectively. River water in Nepal ($n = 14$) showed 57% (8/14), 29% (4/14), and 36% (5/14) detection frequencies of CAstV, MLB-AstV, and VA-AstV, respectively.

Genotype distribution of human AstVs. Genotypes of human AstVs identified in the samples were determined based on phylogenetic analysis (Table 1 and Fig. S1 in the Supplementary Information). As previously reported by other researchers²⁷, types 4 and 8 CAstVs are not distinguishable based on our analysis because their nucleotide sequences are closely related at our target regions, 5'-end of ORF2²⁸. Due to long amplicon sizes after (semi-)nested PCR, sequencing analysis for MLB- and VA-AstVs was conducted based on partial (5'- and 3' ends) sequences of the amplicons. Types 1, 2, 4/8, and 5 CAstVs were identified from all sample groups. Types 3, 6, and 7 were also identified from samples other than groundwater samples with lower frequency. Interestingly, in addition to the CAstV strains, feline AstV (FAstV)-like sequences, which formed a distinct cluster from human AstVs and showed the highest homologies of 89–96% with FAstV based on the BLAST analysis (<https://blast.ncbi.nlm.nih.gov>), were also identified from samples from the U.S. and Nepal. We identified 11 operational taxonomic units (OTUs) that are not classified into clusters with referential sequences.

Only types 1 and 2 MLB-AstVs were identified from the wastewater samples collected in the U.S., whereas all known genotypes of MLB-AstVs (types 1–3) were found from the samples collected in Nepal. Similar to CAstVs, 5 OTUs found in river water samples in Nepal were not classified into clusters with referential sequences.

Types 1–3 VA-AstVs were identified from the wastewater samples collected in the U.S. In addition to types 1–3, type 5 VA-AstVs were identified from river water samples collected in Nepal.

Identification of putative recombinants. Eleven and five OTUs obtained from CAstV in U.S. and MLB-AstV in Nepal samples, respectively, were not classified into clusters with referential sequences (Fig. S1). We assumed that the deviated OTUs are recombinants and carried out Simplot analysis to investigate whether recombination is identified on their genomes. As a result, all the unclassified OTUs were suggested to be recombinant based on the Simplot analysis. Among the putative recombinants, those supported by high bootstrap values (100% among 100 replicates) by the bootscan analysis with two different reference strains at the 5' and 3' ends are shown in Fig. 1 and -2 and others supported by lower bootstrap values (75–100%) are shown in Fig. S2. Putative recombinations of CAstV were observed not only among CAstV strains (type 4/8 and type 1 (Fig. S2A7) and type 4/8 and type 3 (Fig. S2. A8)), but also between CAstV and FAstV strains (type 4/8 or type 2 and FAstV, Fig. 1, Fig. S2A1–6). Regarding MLB-AstV OTUs, all potential recombination events occurred between type 3 MLB-AstV and type 2 or -1 (Figs 2 and S2B).

Seasonal distribution of human AstVs. Table 2 summarizes numbers of OTUs of each AstV clade obtained from wastewater samples (U.S.) in each sampling month. Wastewater samples in U.S. were collected monthly for a year. The frequency of CAstV identification during cooler months, November to April, was significantly higher (96%, 23/24) than that during the rest of the months (42%, 10/24) (chi-square test, $p < 0.01$) (Fig. S1 and Table 2). FAstV-like OTUs were also more frequently identified during cooler months. Similarly, the frequency of VA-AstV identification during the cooler months was higher (50%, 12/24) than that during rest of the months (25%, 6/24), even though the difference was not statistically significant (chi-square test, $p = 0.14$) (Fig. S1 and Table 2). MLB-AstV was identified only in May and June (Fig. S1 and Table 2).

Discussion

In this study, we successfully identified human AstVs, including emerging MLB- and VA-AstVs, in water samples collected from two geographically distinct countries, U.S. and Nepal. On the contrary to its importance as an enteric pathogen, presence of AstVs in water is not as frequently documented as other enteric viruses like noroviruses and rotaviruses⁹. MLB- and VA-AstVs are emerging viruses that were discovered in 2008 and 2009, respectively, and only limited information is available about them^{13,29,30}. Wastewater and environmental water contain viruses excreted from infected individuals including those asymptotically affected¹⁵; thus, investigation of viruses in water can reveal genetic diversity of circulating virus strains in the study area.

Our groundwater samples in Nepal resulted in lower positive ratios of human AstVs than other samples. Previous studies reported that the groundwater samples were less contaminated with other enteric viruses, i.e., human adenoviruses, noroviruses of genogroup I and II, and Aichi virus 1, than the river water samples^{31–33}. All groundwater samples positive for human AstVs were collected from one dug well, which was highly contaminated with waterborne pathogens^{32,33}.

Except for groundwater samples, which showed extremely low human AstV detection frequencies, CAstVs tended to show the highest detection frequencies regardless of the sample types. VA-AstV showed apparently higher detection frequency than MLB-AstV in wastewater samples in U.S., while these two clades of human AstVs were detected in comparable frequencies from river water in Nepal. In our previous study investigating the occurrence of AstVs in wastewater in Japan, MLB-AstV showed higher detection frequency than VA-AstV¹⁷. Some clinical studies compared the abundance of MLB- and VA-AstVs in feces from diarrheal patients, although the number of positive specimens is limited³⁴. In U.S., MLB-AstVs (3/466 positives) were more prevalent than VA-AstVs (1/466 positive) in St. Louis, while only VA-AstVs (1/196 positive) were identified in Seattle^{35,36}. In South Asian countries, only VA-AstVs were identified in Nepal and Pakistan (3/190 and 2/43

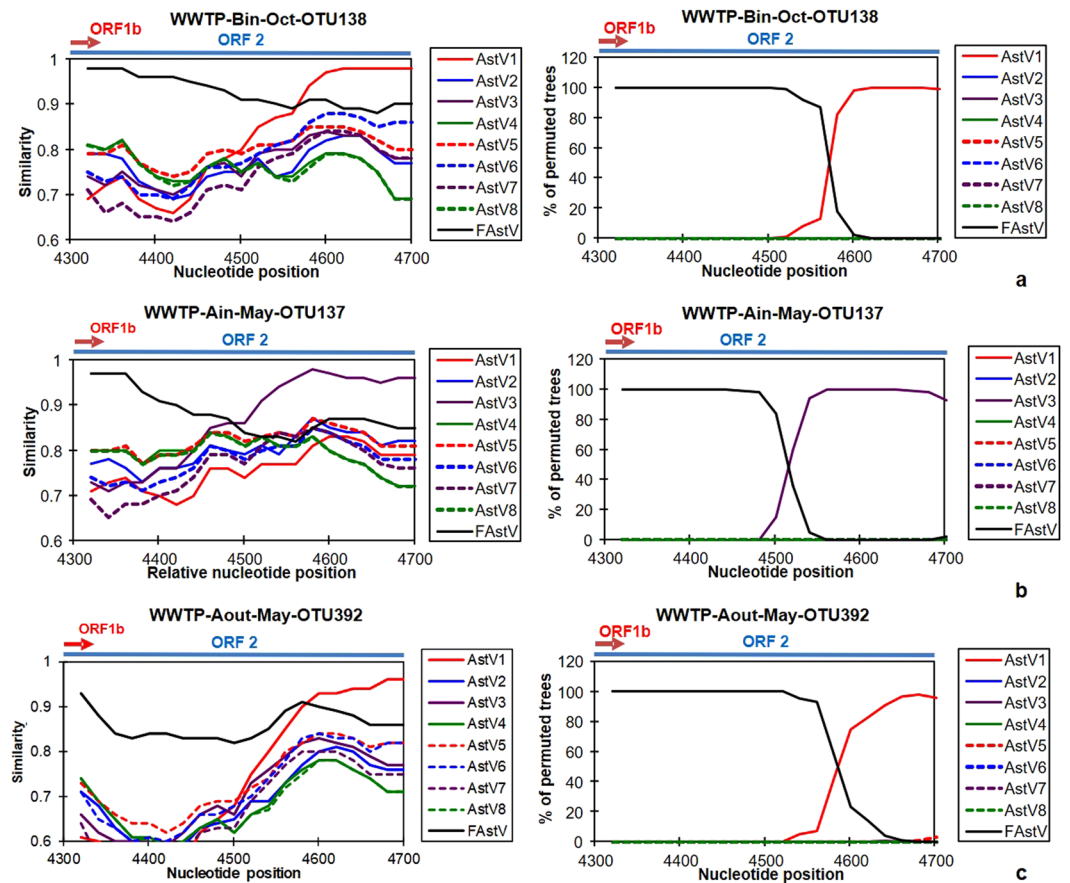


Figure 1. Simplot (left) and bootscan (right) analysis of CAstV OTUs, including FAstV-like OTUs, identified from U.S. samples. OTUs suggested to be recombinants by the phylogenetic analysis (Fig. S1) were subjected to the analysis here. Representative OTUs resulted in 100% bootstrap values among 100 replicates with two different reference strains at the 5' and 3' end in the bootscan analysis are shown here. Results obtained from other OTUs are shown in Fig. S2 in the Supplementary Information. Names of OTUs used as a query sequence in each analysis are shown on upper side of the figures by the following manner, “Origin of the sample (- “type of wastewater”)” - “year or month of sample collection” - “OTU number”. “A” and “B” indicate places of WWTP and “in” and “out” indicate influent and treated effluent, respectively. GenBank accession numbers of referential strains are as follows, CAstV type 1: L23513; -2: L13745; -3: AF141381; -4: DQ344027; -5: DQ028633; -6: GQ495608; -7: AF248738; -8: AF260508; and FAstV: KM017742. Window size and step size for each analysis were set as 100 and 20 nt, respectively. Nucleotide position shown on x-axis corresponds to CAstV type 1 strain Oxford (GenBank accession number: L23513). Arrows and lines indicate locations of ORF1b and ORF2.

positives, respectively), while MLB-AstVs were more prevalent (9/416 positive) than VA-AstVs (3/416 positives) in India^{13,35}. These imply that relative abundance of MLB- and VA-AstVs in water and clinical specimens varies depending on geographical location. It is noteworthy that in all the previous studies in U.S. and South Asia, CAstVs were more predominantly identified than MLB- and VA-AstVs^{13,35,36}.

Diversity of human AstV strains in water samples was studied by Sanger sequencing in previous studies^{17–19,21–23}, except for one study applying next-generation sequencing³⁷. Especially, diversities of MLB- and VA-AstVs in water samples were studied by only one¹⁷ and two studies^{17,18}, respectively. Therefore, only a limited number of genotypes/strains have been identified from water samples. In the present study, we applied next-generation amplicon sequencing and successfully identified diverse CAstV, MLB-, and VA-AstV strains.

CAstV is divided into 8 genotypes based on nucleotide sequence of ORF2¹. In general, types 6 and 7 CAstVs are rarely identified in clinical and water samples^{17,23,37,38}. In this study, we identified these rare genotypes of CAstV even though their detection frequencies were lower than other genotypes. This is probably because of the application of the next-generation sequencing technique, which can read millions of sequences in parallel and therefore enables the identification of minor strains in a sample. These rare genotypes were found from samples collected from both U.S. and Nepal, indicating that rare genotypes of CAstVs are also circulating worldwide. A previous study applied the next-generation amplicon sequencing technique to identify CAstV genotypes in wastewater samples in France³⁷. This previous study identified lower number of genotypes (types 1, 2, 5, and 6) than the present study. One reason for the difference may be PCR assays. The previous study applied single-round PCR with primers MON270 and 269³⁷, which were designed in 1995¹¹. In this study, we applied nested PCR with

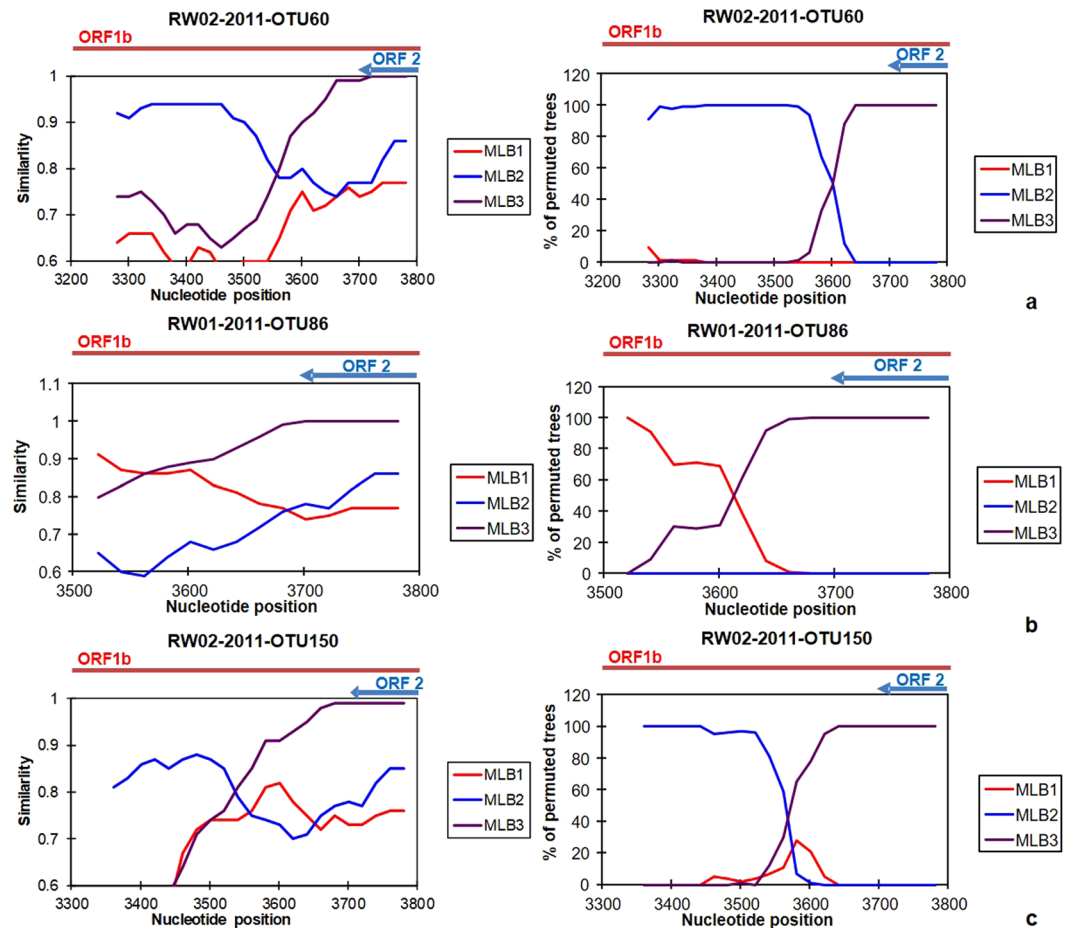


Figure 2. Simplot (left) and bootscan (right) analysis of MLB-AstV OTUs identified from Nepal samples. OTUs suggested to be recombinants by the phylogenetic analysis (Fig. S1) were subjected to the analysis here. Representative OTUs resulted in 100% bootstrap values among 100 replicates with two different reference strains at the 5' and 3' end in the bootscan analysis are shown here. Results obtained from other OTUs are shown in Fig. S2 in the Supplementary Information. Names of OTUs used as a query sequence in each analysis are shown on upper side of the figures by the following manner, “Origin of the sample (- “type of wastewater”)” - “year or month of sample collection” - “OTU number”. GenBank accession numbers of referential strains are as follows, MLB-AstV type 1: FJ222451;-2: JF742759; and -3: JX857870. Window size and step size for each analysis were set as 100 and 20 nt, respectively. Nucleotide position shown on x-axis corresponds to MLB-AstV type1 (GenBank accession number: FJ222451). Arrows and lines indicate locations of ORF1b and ORF2.

primers designed in 2014 based on updated list of AstV strains including contemporary circulating ones²⁷. Thus, our assay should be more suitable in identifying broad range of CAstVs sensitively.

Not only CAstV but also FAstV-like sequences were identified in our investigation. FAstVs (mammalian AstV-2) are more closely related to CAstVs (mammalian AstV-1) than any other AstV species¹. Accordingly, our reverse transcription (RT)-nested PCR primers showed only 3-base mismatches with a FAstV sequence in the database (GenBank, KF499111), and therefore, it is highly possible that the assay amplifies FAstV genomes as well. A previous study employing another set of primers for detection of CAstV also found FAstV-like sequence from surface water in urbanized area of Singapore²³. Unexpected amplification of FAstV gene by primers targeting CAstV may be a common issue and may lead to overestimation of the presence of CAstVs, especially in studies investigating the effect of human fecal contamination of environmental samples. In accordance with previous studies targeting both clinical and environmental samples^{1,17,37,39}, CAstV became prevalent during cooler months in this study. FAstV-like sequences also followed the same seasonal pattern, although the number of positive sample is too small to conclude. This suggests that FAstV is closely related to CAstV in terms of genetic distance and seasonal pattern. In a previous study, FAstV has been detected from a domestic cat⁴⁰. Besides, another previous study has revealed that canine kobuvirus is present in municipal wastewater⁴¹. These suggest that viruses excreted from pet animals can contaminate water environment with human viruses. It is possible that cats in the study area are the source of FAstV-like sequences identified in this study.

Two previous studies have investigated the presence of MLB-AstVs in water samples^{17,18}. Each study identified one genotype, i.e., one study identified type 2¹⁷ and the other identified type 1¹⁸. In the present study, multiple types of MLB-AstVs were identified. All three types of MLB-AstVs were found from samples in Nepal, while

	WWTP	2011					2012						
		Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July
CAstV	A-in ^b			14	56	14	58	32	116	27	5	4	
	A-out				72	72		44	18	55	167	4	
	B-in	1	1		57	36	29	17	33	18		7	
	B-out		36		16	20	6	8	30	6	10		
MLB-AstV ^a	A-in											11–29	
	A-out												
	B-in												
	B-out										24–11		
VA-AstV	A-in				1-1			9-2	24-16	21-20	30-21	11-7	43-17
	A-out							16-14		13-21			
	B-in				3-1	8-3	7-3	12-13		36-27	12-13		
	B-out			1-2	1-0						8-9		

Table 2. Number of OTUs of each human AstV clade determined from wastewater samples in U.S. ^aNumber of MLB- and VA-AstVs OTUs were shown as follows, “number of OTUs formed from 5′- end of amplicons sequences – those formed from 3′-end of amplicons sequences”. ^b“A” and “B” indicate places of WWTP. “in” and “out” indicate influent and treated effluent, respectively.

type 3 was not found from wastewater in U.S. Our findings suggest that the genotype distribution of MLB-AstVs depends on geographical regions. In our wastewater samples, MLB-AstVs were identified only in early summer (May and June). On the contrary, our previous study suggested that MLB-AstV become prevalent during winter in wastewater in Japan¹⁷. This implies that seasonal pattern of MLB-AstVs varies depending on geographical regions, although it is not conclusive. It is important to note that the previous study resulted in extremely high detection frequency probably because the study investigated samples collected during only cooler months¹⁷.

To the best of our knowledge, this is the first study identifying types 3 and 5 VA-AstVs in water samples. Type 5 VA-AstV is a newly identified type, discovered from a pediatric stool sample from Gambia⁵. Our result demonstrates that type 5 VA-AstVs are also circulating in Nepal. In a previous study, type 4 VA-AstV was identified from stool samples in Nepal⁴¹; however, this genotype was not identified in the present study. Analysis of wastewater samples suggested that seasonal pattern of VA-AstV is similar to that of CAstV, become prevalent during colder month, but the epidemic period seemed to be shorter than that of CAstV. Such a seasonal pattern of VA-AstV is not conclusive. Future establishment of quantitative assay for VA-AstV and long-term monitoring should provide more insights.

It has been suggested that recombination events play an important role in the evolution of AstVs⁴³. We identified some potential recombinant strains in the present study. Our RT-nested PCR assay targeting the ORF1b-ORF2 junction region, which is known as a “hotspot” of recombination, allowed us to identify potential recombinant strains^{43,44}. Some CAstV genomes were considered potential recombinants. The most frequently reported recombination breakpoints are nucleotide positions between 4100 and 4400 in CAstV-1 (GenBank accession number: L23513), just around the ORF1b-ORF2 junction region^{43–46}. Most of CAstV recombination breakpoints suggested by our analysis are nucleotide positions between 4500 and 4600, which are downstream of those frequently reported. To our knowledge, recombination events at the area were not reported in previous studies but those at positions apart from the junction region are possible^{43,47}.

Interestingly, recombination between human and FAsTV was also identified. The recombination between human and feline strains suggests zoonotic transmission of AstVs, which may lead to an emerging and/or unrecognized risk. Indeed, CAstVs have been detected from non-human primates and piglet samples^{48,49}, and evidences of recombination between human- and these animal AstVs were concurrently reported^{48,49}. Future investigation of zoonotic potential of AstVs is required. Recombination within MLB-AstV strains, which has not been reported in previous studies, was also suggested in this study. Our findings can prompt future investigations of recombination events occurring among broad range of AstVs.

Water samples impacted by wastewater potentially contain a variety of viruses originated from a wide number of people. Next-generation sequencing is an ideal technology to identify diverse viruses in a water sample because of its ability for massive parallel sequencing. In the present study, we successfully determined diverse AstV sequences including rare types (types 6 and 7 CAstVs), an emerging type (type 5 VA-AstV), and recombinants by applying next-generation amplicon sequencing approach. The recombinant of CAstV and FAsTV strongly suggests zoonotic potential of this AstV strain. The present study also demonstrated that genetically diverse AstVs are circulating in the studied geographical regions.

Methods

Sample collection and concentration. Influent (n = 24) and effluent (n = 24) wastewater samples were collected from two wastewater treatment plants (WWTPs, WWTP-A and -B) in Arizona, U.S. monthly between August 2011 and July 2012, as described in Kitajima *et al.*⁵⁰. One of the WWTPs utilized a conventional activated sludge process and the other utilized a biological tricking filter process. An adsorption-elution method followed by ultrafiltration⁵¹ was used to concentrate 100 and 1,000 mL of influent and treated effluent, respectively, to obtain a final volume of approximately 650 μL.

Groundwater (GW, $n = 37$) and river water (RW, $n = 14$) samples were collected from 15 and 8 sites, respectively, between August 2009 and May 2011 in the Kathmandu Valley, Nepal as described previously^{31–33}. An electronegative membrane vortex method⁵² was applied to ground and river water samples to obtain 12 mL of virus concentrates. The volumes filtered were 50 or 100 mL for river water and 1,000 mL for groundwater samples except for one sample that allowed filtration of only 50 mL.

RNA extraction and RT-(semi-)nested-PCR. Viral RNA in the virus concentrates obtained in U.S. and Nepal was extracted using a ZR Viral DNA/RNA Kit (Zymo Research, Irvine, CA, USA) and QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), respectively. Extracted RNA was subjected to RT using a High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. Subsequently, (semi-)nested PCR assays targeting the ORF1b-ORF2 junction regions of CAsV, MLB-AstV, and VA-AstV were performed separately as described previously¹⁷ (Table S2 in the Supplementary Information). Amplification of each target gene was confirmed by visualization under a UV lamp after electrophoresis in a 1.5% agarose gel stained by GelRed™ (Wako, Osaka, Japan). In the first-round PCR, primers AHAsVF1 and AHAsVR1, SF0073 and AHMLBR1, and AHVAF1 and AHVAR1 were used to amplify CAsV, MLB-, and VA-AstV genes, respectively. In the second-round PCR, primers AHAsVF2 and AHAsVR2, F0073 and AHMLBR2, and AHVAF2 and AHVAR2 were used to amplify the first PCR amplicons of CAsV, MLB-, and VA-AstV genes, respectively. Resultant second-round PCR amplicon sizes of CAsV, MLB-AstV, and VA-AstV genes were expected to be 407, 689, and 663 bp, respectively.

Next-generation amplicon sequencing. The second-round PCR amplicons with expected size for each clade of human AstVs were sequenced with the Illumina MiSeq platform (Illumina, San Diego, CA, USA). Briefly, a sequencing reaction mixture was prepared using the TruSeq DNA LT sample preparation kit (Illumina), and loaded on a MiSeq Reagent Kit v3 (600 cycles) (Illumina), according to the manufacturer's instructions, which allows the system to obtain read length of up to 300 bp. Nucleotide sequencing was performed for 301 cycles for both ends. Obtained sequencing data were trimmed using CLC Genomics Workbench 7.0 software (CLC bio, Aarhus, Denmark). The trimmed sequence reads obtained from CAsV were paired. Those obtained from MLB- and VA-AstVs were not paired because their expected amplicon sizes (689 and 663 bp, respectively) were longer than the length that can be read by the system. Then, edited sequence reads that showed 70% or higher nucleotide identities to the reference CAsV, MLB-AstV, or VA-AstV strains (Oxford strain: GenBank acc. no. L23513; MLB1 strain: FJ222451; and VA1 strain FJ973620, respectively) were extracted using "Map Reads to Reference" command of the CLC Genomics Workbench 7.0 software and the extracted sequences were subjected to further analysis. The trimmed sequence reads sharing 97% or higher nucleotide sequence identities were assigned in an OTU. Representative sequences from each OTU were used for phylogenetic analysis and genotyping. As mentioned above, sequence reads obtained from MLB- and VA-AstVs were not paired due to their long amplicon sizes. Thus, phylogenetic analysis for MLB- and VA-AstVs were conducted based on partial (5'- and 3' ends) sequences of the amplicons. OTUs that were not classified within any known phylogenetic clades were analyzed for recombination with the Simplot software version 3.5.1⁵³ (<http://sray.med.som.jhmi.edu/SCSoftware/simplot/>). For phylogenetic analysis, OTUs were aligned using Clustal W program version 1.83 (<http://clustalw.ddbj.nig.ac.jp/top-e.html>). The distances were calculated by Kimura's two-parameter method⁵⁴ and phylogenetic dendrograms from bootstrap analysis with 1,000 replicates were generated by the neighbor joining-method.

Statistical analysis. Chi-square test was performed to determine seasonality of human AstVs in wastewater samples in U.S. using R software version 3.4.0 (<https://www.r-project.org/>). *P*-values of < 0.01 were considered statistically significant.

References

- Bosch, A., Pintó, R. M. & Guix, S. Human astroviruses. *Clin. Microbiol. Rev.* **27**(4), 1048–1074 (2014).
- Andreas, M. S. *et al.* Adenovirus, calicivirus and astrovirus detection in fecal samples of hospitalized children with acute gastroenteritis from Campo Grande, MS, Brazil. *Mem. Inst. Oswaldo Cruz.* **103**(7), 741–744 (2008).
- Dai, Y. C. *et al.* Development of real-time and nested RT-PCR to detect astrovirus and one-year survey of astrovirus in Jiangmen City, China. *Arch. Virol.* **155**(6), 977–982 (2010).
- Khamrin, P. *et al.* A single-tube multiplex PCR for rapid detection in feces of 10 viruses causing diarrhea. *J. Virol. Methods.* **173**(2), 390–393 (2011).
- Meyer, C. T. *et al.* Prevalence of classic, MLB-clade and VA-clade Astroviruses in Kenya and The Gambia. *Virol. J.* **12**, 78 (2015).
- Quan, P. L. *et al.* Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerg. Infect. Dis.* **16**(6), 918–925 (2010).
- Lum, S. *et al.* An emerging opportunistic infection: fatal astrovirus (VA1/HMO-C) encephalitis in a pediatric stem cell transplant recipient. *Transpl. Infect. Dis.* **18**(6), 960–964 (2016).
- Sato, M. *et al.* Acute encephalopathy in an immunocompromised boy with astrovirus-MLB1 infection detected by next generation sequencing. *J. Clin. Virol.* **78**, 66–70 (2016).
- Yang, B. Y. *et al.* Rapid and sensitive detection of human astrovirus in water samples by loop-mediated isothermal amplification with hydroxynaphthol blue dye. *BMC Microbiol.* **14**, <https://doi.org/10.1186/1471-2180-14-38> (2014).
- Haramoto, E. *et al.* A review on recent progress in the detection methods and prevalence of human enteric viruses in water. *Water Res.* **135**, 168–186 (2018).
- Noel, J. S., Lee, T. W., Kurtz, J. B., Glass, R. I. & Monroe, S. S. Typing of human astroviruses from clinical isolates by enzyme immunoassay and nucleotide sequencing. *J. Clin. Microbiol.* **33**(4), 797–801 (1995).
- Finkbeiner, S. R., Le, B. M., Holtz, L. R., Storch, G. A. & Wang, D. Detection of newly described astrovirus MLB1 in stool samples from children. *Emerg. Infect. Dis.* **15**(3), 441–444 (2009a).
- Kapoor, A. *et al.* Multiple novel astrovirus species in human stool. *J. Gen. Virol.* **90**(12), 2965–2972 (2009).
- Méndez-Toss, M. *et al.* Prevalence and genetic diversity of human astroviruses in Mexican children with symptomatic and asymptomatic infections. *J. Clin. Microbiol.* **42**(1), 151–157 (2004).
- Lopman, B. *et al.* Environmental transmission of norovirus gastroenteritis. *Curr. Opin. Virol.* **2**(1), 96–102 (2012).

16. Pinto, R. M., Abad, F. X., Gajardo, R. & Bosch, A. Detection of infectious astroviruses in water. *Appl. Environ. Microbiol.* **62**(5), 1811–1813 (1996).
17. Hata, A., Katayama, H., Kitajima, M. & Furumai, H. Wastewater Analysis Indicates that Genetically Diverse Astroviruses, Including Strains Belonging to Novel Clades MLB and VA, Are Circulating within Japanese Populations. *Appl. Environ. Microbiol.* **81**(15), 4932–4939 (2015).
18. Lizasoain, A. *et al.* Environmental assessment reveals the presence of MLB-1 human astrovirus in Uruguay. *J Appl Microbiol.* **119**(3), 859–867 (2015).
19. Yokoi, H., Kitahashi, T., Tanaka, T. & Utagawa, E. Detection of astrovirus RNA from sewage works, seawater and native oysters samples in Chiba City, Japan using reverse transcription-polymerase chain reaction. *Kansenshogaku Zasshi.* **75**, 263–269 (2001). (In Japanese).
20. Hansman, G. S. *et al.* Detection of human enteric viruses in Japanese clams. *J. Food. Prot.* **71**(8), 1689–1695 (2008).
21. Rodríguez-Díaz, J. *et al.* Detection and characterization of waterborne gastroenteritis viruses in urban sewage and sewage-polluted river waters in Caracas, Venezuela. *Appl. Environ. Microbiol.* **75**(2), 387–394 (2009).
22. Aw, T. G. & Gin, K. Y. Environmental surveillance and molecular characterization of human enteric viruses in tropical urban wastewaters. *J. Appl. Microbiol.* **109**(2), 716–730 (2010).
23. Aw, T. G. & Gin, K. Y. Prevalence and genetic diversity of waterborne pathogenic viruses in surface waters of tropical urban catchments. *J. Appl. Microbiol.* **110**(4), 903–914 (2011).
24. van den Berg, H., Lodder, W., van der Poel, W., Vennema, H. & de Roda Husman, A. M. Genetic diversity of noroviruses in raw and treated sewage water. *Res. Microbiol.* **156**(4), 532–540 (2005).
25. Kitajima, M. *et al.* Detection and genetic analysis of human sapoviruses in river water in Japan. *Appl. Environ. Microbiol.* **76**(8), 2461–2467 (2010).
26. Kitajima, M., Haramoto, E., Phanuwat, C. & Katayama, H. Prevalence and genetic diversity of Aichi viruses in wastewater and river water in Japan. *Appl. Environ. Microbiol.* **77**(6), 2184–2217 (2011).
27. Hata, A., Kitajima, M., Tajiri-Utagawa, E. & Katayama, H. Development of a high resolution melting analysis for detection and differentiation of human astroviruses. *J. Virol. Methods.* **200**, 29–34 (2014).
28. Méndez-Toss, M., Romero-Guido, P., Munguía, M. E., Méndez, E. & Arias, C. F. Molecular analysis of a serotype 8 human astrovirus genome. *J. Gen. Virol.* **81**(Pt 12), 2891–2897 (2000).
29. Finkbeiner, S. R. *et al.* Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog.* **4**(2), <https://doi.org/10.1371/journal.ppat.1000011> (2008).
30. Finkbeiner, S. R. *et al.* Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. *J. Virol.* **83**(20), 10836–10839 (2009b).
31. Haramoto, E., Yamada, K. & Nishida, K. Prevalence of protozoa, viruses, coliphages and indicator bacteria in groundwater and river water in the Kathmandu Valley, Nepal. *Trans R Soc Trop Med Hyg.* **105**(12), 711–716 (2011).
32. Haramoto, E. & Kitajima, M. Quantification and Genotyping of Aichi Virus 1 in Water Samples in the Kathmandu Valley, Nepal. *Food Environ. Virol.* **9**(3), 350–353 (2017).
33. Haramoto, E. Detection of waterborne protozoa, viruses, and bacteria in groundwater and other water samples in the Kathmandu Valley, Nepal. *IOP Conf. Ser.: Earth Environ. Sci.* **120**(1), <https://doi.org/10.1088/1755-1315/120/1/012004> (2018).
34. Vu, D. L., Cordey, S., Brito, F. & Kaiser, L. Novel human astroviruses: Novel human diseases? *J. Clin. Virol.* **82**, 56–63 (2016).
35. Finkbeiner, S. R. *et al.* Human stool contains a previously unrecognized diversity of novel astroviruses. *Virol. J.* **6**, 161 (2009c).
36. Melamed, R. *et al.* Case-Control Assessment of the Roles of Noroviruses, Human Bocaviruses 2, 3, and 4, and Novel Polyomaviruses and Astroviruses in Acute Childhood Diarrhea. *J. Pediatric. Infect. Dis. Soc.* **6**(3), e49–e54 (2017).
37. Prevost, B. *et al.* Deciphering the Diversities of Astroviruses and Noroviruses in Wastewater Treatment Plant Effluents by a High-Throughput Sequencing Method. *Appl. Environ. Microbiol.* **81**(20), 7215–7222 (2015).
38. Morsy El-Senousy, W., Guix, S., Abid, I., Pintó, R. M. & Bosch, A. Removal of astrovirus from water and sewage treatment plants, evaluated by a competitive reverse transcription-PCR. *Appl. Environ. Microbiol.* **73**(1), 164–167 (2007).
39. Guimarães, F. R. *et al.* Molecular detection of human astrovirus in an urban sewage treatment plant in Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo. Cruz.* **103**(8), 819–823 (2008).
40. Lau, S. K. *et al.* Complete genome sequence of a novel feline astrovirus from a domestic cat in Hong Kong. *Genome Announc.* **1**(5), <https://doi.org/10.1128/genomeA.00708-13> (2013).
41. Yamashita, T. *et al.* Molecular detection and nucleotide sequence analysis of a new Aichi virus closely related to canine kobuvirus in sewage samples. *J. Med. Microbiol.* **63**(5), 715–720 (2014).
42. Jiang, H. *et al.* Comparison of novel MLB-clade, VA-clade and classic human astroviruses highlights constrained evolution of the classic human astrovirus nonstructural genes. *Virology.* **436**(1), 8–14 (2013).
43. Babkin, I. V., Tikunov, A. Y., Sedelnikova, D. A., Zhirakovskaia, E. V. & Tikunova, N. V. Recombination analysis based on the HAsTV-2 and HAsTV-4 complete genomes. *Infect. Genet. Evol.* **22**, 94–102 (2014).
44. Walter, J. E. *et al.* Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children. *Arch. Virol.* **146**(12), 2357–2367 (2001).
45. Wolfardt, M., Kiulia, N. M., Mwenda, J. M. & Taylor, M. B. Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. *J. Clin. Microbiol.* **49**(2), 728–731 (2011).
46. De Grazia, S. *et al.* Genetic heterogeneity and recombination in human type 2 astroviruses. *J. Clin. Microbiol.* **50**(11), 3760–3764 (2012).
47. Babkin, I. V., Tikunov, A. Y., Zhirakovskaia, E. V., Netesov, S. V. & Tikunova, N. V. High evolutionary rate of human astrovirus. *Infect. Genet. Evol.* **12**(2), 435–442 (2012).
48. Ulloa, J. C. & Gutiérrez, M. F. Genomic analysis of two ORF2 segments of new porcine astrovirus isolates and their close relationship with human astroviruses. *Can. J. Microbiol.* **56**(7), 569–577 (2010).
49. Karlsson, E. A. *et al.* Non-Human Primates Harbor Diverse Mammalian and Avian Astroviruses Including Those Associated with Human Infections. *PLoS Pathog.* **11**(11), e1005225 (2015).
50. Kitajima, M., Iker, B. C., Pepper, I. L. & Gerba, C. P. Relative abundance and treatment reduction of viruses during wastewater treatment processes—identification of potential viral indicators. *Sci. Total. Environ.* **488–489**, 290–296 (2014).
51. Katayama, H., Shimasaki, A. & Ohgaki, S. Development of a virus concentration method and its application to detection of enterovirus and norwalk virus from coastal seawater. *Appl. Environ. Microbiol.* **68**(3), 1033–1039 (2002).
52. Haramoto, E., Katayama, H., Asami, M. & Akiba, M. Development of a novel method for simultaneous concentration of viruses and protozoa from a single water sample. *J. Virol. Methods.* **182**(1–2), 62–69 (2012).
53. Lole, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**(1), 152–160 (1999).
54. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**(2), 111–120 (1980).

Acknowledgements

This work was supported by grant-in-aid for JSPS Fellows (grant number: 14J03643) and JSPS KAKENHI (grant numbers: JP 26820222, JP 26289182, and 17H03332) from the Japan Society for the Promotion of Science.

Author Contributions

A.H. analyzed samples and wrote the initial draft of the manuscript. E.H., M.K. and C.G. collected samples. S.L., M.I., and H.T. contributed to the analysis and assisted in the preparation of the manuscript. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30217-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018