Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children

J. E. Walter1,3, J. Briggs1, M. L. Guerrero2, D. O. Matson1, L. K. Pickering1, G. Ruiz-Palacios2, T. Berke1, and D. K. Mitchell1

1Center for Pediatric Research, Eastern Virginia Medical School, Children’s Hospital of The King’s Daughters, Norfolk, Virginia, USA
2Instituto Nacional de la Nutrición, Mexico City, Mexico
3Department of Clinical Chemistry, Pecs University Medical School, Hungary

Accepted July 5, 2001

Summary. We report a naturally occurring human astrovirus (HAstV) strain detected in two different geographic locations. We identified two isolates of this strain in a diarrhea outbreak at a child care center in Houston, Texas; and two isolates in diarrhea stool samples from two children in Mexico City. All four isolates were detected in stool samples by enzyme immunoassay (EIA). One of the Mexican isolates was typed by EIA and all four isolates were HAstV-5 by typing RT-PCR. The four isolates were >97% nucleotide-identical in two different genomic regions: ORF1a (246nt), and the 3′ end of the genome (471nt). One isolate from each geographic location was further sequenced in the transition region from ORF1b to ORF2 (1255nt) and this region of the two isolates showed ≥99% nt identity. Phylogenetic analyses of sequences of eight HAstV antigenic types and the novel strain in the transition region demonstrated the new strain being closely related to HAstV-3 in ORF1b, but closest to HAstV-5 in ORF2. These results and high sequence identity among all HAstV antigenic types in the transition region and RNA structural predictions supported a potential recombination site at the ORF1b/ORF2 junction. This is the first evidence that recombination occurs among human astroviruses.

Introduction

Astroviruses (AstVs) contain a single-stranded, positive-sense RNA (+ssRNA) genome approximately 6,800 nucleotides (nt) long. The genome includes three open reading frames: ORF1a, ORF1b, and ORF2 (Fig. 1) [6, 13]. ORF1a and 1b encode non-structural proteins, including a serine protease (Pro) and RNA-dependent RNA polymerase (Pol), respectively, and these regions contain
sequences highly conserved among AstVs. ORF2 encodes the capsid protein and is highly variable at the 3′ end.

The family Astroviridae is distinguished from other virus families that are not enveloped and have ssRNA genomes, such as Picornaviridae and Caliciviridae, by the presence of ribosomal frameshifting between ORF1a and ORF1b and the lack of the helicase domain [6]. However, AstVs contain eight characteristic amino acid motifs in the Pol region (e.g. YGDD) typical of +ssRNA virus polymerases, indicating that astrovirus Pol belongs to supergroup I polymerases, which includes the Pol enzyme of picornaviruses, caliciviruses, potyviruses and several other plant viruses [6]. +ssRNA plant viruses also share similarities with human astroviruses (HAstVs), such as the substitution of the catalytic cysteine with serine and the genomic organization resembles that of Luteoviridae [24].

Belliot et al. characterized three different regions of the astrovirus genome (ORF1a, ORF1b and 5′ end of ORF2) of 64 HAstV isolates from seven countries [2]. Phylogenetic analyses of these partial ORF sequences suggested that genotypes could be assigned equally well by analysis of any of these three regions because of high correlation among the regions. Their findings would suggest that AstV recombinants did not occur among their characterized strains.

The aims of this study were 1) to characterize four HAstV isolates whose genotyping, sequotyping and antigenic typing results for the polymerase (ORF1b) and capsid (ORF2) genes resulted in conflicting classification and 2) to determine their genetic and antigenic relationships to known HAstV serotypes.

**Materials and methods**

**Stool specimens**

HAstVs were detected in two separate studies. In the first study, stool specimens were collected from 1989 to 1992 during diarrhea outbreaks among children attending childcare centers in Houston, Texas [12]. In the second study, 214 children in a periurban community of Mexico City were monitored prospectively for diarrhea from birth to 18 months of age from 1988 to 1991 [5]. All collected stool specimens were stored at −70 °C until HAstV testing.

**Primers**

Primer pair Mon2 [12] /DM4 (5′ CTA CAG TTC ACT CAA ATG AA 3′, 5531-5550nt in HAstV-1 [Acc# L23513]) was used to amplify a 1.2 kb segment of the 3′ end of the genome.
Primer pair Mon270/Mon344 [2, 21] was used to amplify a 1.2 kb segment of the transition region overlapping ORF1b and ORF2. Internal primers DM17 (5′ GGTTTTGGT CCT GTG ACA CC 3′; location 4563-4544nt in HAstV-1 [Acc# L23513]) and DM15 (5′ GATAGG CTC TCC ACT ACA CC 3′, location 3901-3920nt in HAstV-2 [Acc# L13745]) were used to amplify a ~600 bp product from the ORF1b-ORF2 transition region of reference HAstV strains. The novel astrovirus was also genotyped by RT-PCR using Mon2/AST-S5 [11].

Detection of HAstVs

Stool samples from Houston and Mexico obtained during diarrhea episodes were tested for HAstV by EIA [5] and RT-PCR. For RT-PCR, viral RNA was extracted from 300 μL of stool suspension (10–50% in water) by a Genetron™-TRIZOL™ method as previously described [23]. Amplicons were generated by RT-PCR utilizing a type-common primer pair, Mon348/Mon340, that targets a region of ORF1a upstream from the Pro motif and that allows detection of all eight serotypes [2].

Typing of HAstVs

HAstV-positive specimens from Mexico were typed by EIA [5] and genotyped by RT-PCR [23]. All typing RT-PCR reagents and conditions were identical to those for type-common detection except the annealing temperature was 40 °C (Mon2/AST-S5).

RT-PCR in the Orf1b-ORF2 region

All RT-PCR reagents and conditions were identical to those for type-common detection except the annealing temperature was 40 °C (Mon270/Mon344, DM17/DM15).

Sequence analysis

RT-PCR amplicons of the four strains were cloned into pGEM-T (Promega, Madison, WI) and sequenced in the ALFwin automated sequencer system (Amersham Pharmacia Biotech, Uppsala, Sweden) using fluorescence-labeled primers and SequiTherm Excel™ II Long-Read DNA Sequencing Kit-ALF™ (Epicentre Technologies, Madison, WI), according to the manufacturer’s recommendations.

Published HAstV capsid gene sequences used in the phylogenetic analyses included HAstV-1 [L23513], HAstV-2 [L13745], HAstV-3 [AF117209], HAstV-4 [Z33883], HAstV-5 [U15136], HAstV-6 [Z46658]. The capsid gene of a HAstV-7 Oxford reference strain was sequenced in our laboratory [AF248738]. The 1.2 kb amplicon of the 3′ end of the genome of H2067 was sequenced and submitted to GenBank [AF292080]. Amplicons (1.2 kb) from the junction region between ORF1b and ORF2 for HAstV-3 to 7 Oxford strains [AF292074-8], HAstV-8 South African strain (As20) [AF292073], and isolate H2067 [AF292072] were sequenced and submitted to GenBank.

Sequence verification and analysis was performed using OMIGA (v2.0, Genetics Computer Group, Madison, WI). ClustalX was used to create multiple alignments of the amino acid sequences of selected isolates [22]. Nucleic acid sequences were added and aligned in GeneDoc v2.3 using the corresponding amino acid alignment as template, resulting in a consensus length of 1264nt in the junction region between ORF1b and ORF2 [20].

Phylogenetic trees were constructed from the nucleic acid sequence alignments using the maximum likelihood algorithm of the program DNAML of PHYLIP (v 3.52c) running in UNIX environment [3]. The global rearrangement option was invoked and the order of the sequence input was randomized one hundred times.

Pairwise identity was calculated for reference types and H2067 in the transition region of ORF1b-ORF2 based on the original alignment. Identities between selected reference types
and H2067 were visualized by dot plot comparison. The parameters for the dot plot were the following: window size: 25; jump size: 1; hash value: 1; and stringency: 0.85.

The nucleic acid sequence of the junction region between ORF1b and ORF2 was further analyzed to characterize the possible recombination site: the length, extent of identity and AU-GC content among all human reference strains and one animal strain (porcine [AB037272]) were evaluated. Secondary RNA structure was computed using Zuker’s RNA secondary structure predicting program (http://www.ibc.wustl.edu/~zuker/). This structure analysis was utilized to predict local secondary structure elements that may represent recombination hotspots.

Results

Detection of viral antigen by EIA and RT-PCR

Two astrovirus strains (H2002, H2067) were detected in diarrhea stool samples collected from two children during a HAstV outbreak in May 1990 at a childcare center in Houston, Texas. The other two strains (M240, M370) were detected in stool samples of diarrhea episodes in June 1990 and April 1991 from two children residing in Mexico City. All four HAstV isolates were detected by type-common EIA [5].

Antigenic and genomic typing

Typing EIA identified M370 as HAstV-5 and M240 was untypeable [5]. EIA typing was not available for H2002 and H2067. The inability to EIA type M240 was attributed to a low titer of virus in the stool specimen, which was further suggested by the weak amplicon signal after RT-PCR amplification. The isolates from Houston were not typed by EIA. All four isolates were genotyped as HAstV-5 using Mon2/AST-S5, type-specific primers targeting the 3′ end of ORF2, and the type assignments were confirmed by sequence analysis (sequence analysis data not shown) [11]. A longer amplicon of H2067 (1128bp) using primers Mon2/DM4 was sequenced.

Sequence comparison

The four isolates were 97–100% identical in the ORF1a region (246nt) and 98–100% identical at the 3′ end of ORF2 (365nt). Amplicons (1255nt) were obtained for two of the four isolates (H2067 and M370) in the ORF1b/ORF2 transition region and were 99% identical to each other. Pairwise sequence comparison of 1128nt at the 3′ end of ORF2 (Mon2/DM4) of H2067 confirmed it’s being closest (94%) to HAstV-5.

Sequence analysis of the transition region of ORF1b-ORF2 (1264 nt consensus length) is shown in Fig. 2. The overlapping clone was divided into two regions (5′ end 576nt, 3′ end 637nt) by a highly identical 52nt region. These two regions represented two different ORFs: ORF1b was represented in the 5′ end (576 nt) and ORF2 was represented in the 3′ end (637nt) of this clone.

Pairwise comparison of nucleotide and amino acid sequences of all reference types with H2067 is shown in Table 1. In ORF1b (576nt), H2067 is closer to HAstV-3 (92%nt, 98%aa) than to HAstV-5 (88%nt, 96%aa). In the 5′ end of
Comparative sequence analysis of the reference human astrovirus type 3 and 5 and H2067 [AF292080] in the transition region of ORF1b-ORF2 (1255nt, x axis). The transition region was divided into two regions (5′ end 576nt, 3′ end 637nt), excluding a 52nt highly identical region in the middle. Dotted lines indicate the hypothesized 120nt-long recombination site. Maximum likelihood phylograms of 576nt of ORF1b and 637nt of ORF2 5′ end are in the left upper and right lower corners, respectively. Phylogenetic distances are given as number of substitutions per site. Identities between selected strains (T5, T3, and H2067) are presented on the diagonal created by dot plot analysis. A high rate of identity is visualized by a dark continuous line.

ORF2 (637nt), H2067 is closer to HAstV-5 (96%nt, 99%aa) than to HAstV-3 (81%nt, 93%). Other HAstV types were less identical to H2067 in the ORF1b (84–90%nt, 94–97%aa) and ORF2 (80–83%nt, 89–92%aa) regions.

Pairwise identities between selected strains (T5, T3, and H2067) are presented by dot plot, in which a higher rate of identity is represented by continuous dots (Fig. 2, diagonal). The dot plot results demonstrated greater identity between H2067 and HAstV-3 in ORF1b and greater identity between H2067 and HAstV-5 at the 5′ end of ORF2. Comparison of HAstV-3 and 5 in this transition region provided the baseline for analysis.

Phylogenetic analysis

A maximum likelihood phylogram for the 576nt region of ORF1b is in the left upper corner of Fig. 2 and shows that H2067 was closest to HAstV-3 and distinct
Table 1. Pairwise nucleotide (nt) and amino acid (aa) identities (%) of H2067 and all HAstV reference types in the junction region of ORF1b (576nt) (A) and ORF2 (637nt) (B). Comparison of T3 and T5 to H2067 are highlighted from HAstV-5. The phylogram of the 637nt portion of ORF2 shows that H2067 was closest to HAstV-5 and distinct from HAstV-3 (Fig. 2, lower right tree).

RNA structure analysis

A 120nt region at the junction of ORF1b and ORF2 was selected for further sequence analysis to characterize the putative recombination site. The consensus nucleotide sequence of the region is shown in Fig. 3. All human strains are 93–98% identical in this region. This 120nt region incorporates a 52nt region where sequence identity is 99–100% among human strains.

Figure 3 demonstrates the predicted RNA secondary structure for H2067 in the 120nt junction region between ORF1b and ORF2. The predicted structure has a triple-hairpin conformation. The predicted RNA secondary structure was identical among the analyzed human strains. Although the porcine astrovirus has only 80–85% nt identity in this 120nt region, secondary structure analysis predicted the same triple-hairpin conformation (data not shown). The triple-hairpin structure includes the start codon for the capsid gene, as well as the stop codon for the Pol gene.
Recombinant human astrovirus

Fig. 3. RNA secondary structure of the hypothesized 120nt recombination site in H2067 (520nt–639nt). Nucleotides are numbered according to their position in the 1255nt RT-PCR product [AF292080], including the junction of ORF1b and ORF2. In the hairpin structure, the location for the start codon of ORF2 and the stop codon for ORF1b are indicated. At the bottom of the figure the consensus sequence of the corresponding regions of all reference human strains is described. Uppercase letters indicate 100% and lower case letters indicate ≥89% identity. Dots represent divergence in the nucleotide sequence of the eight reference types. The highly conserved 52nt region is underlined. In the consensus sequence, the start codon for ORF2 and the stop codon for ORF1b are highlighted.

Discussion

In this study we characterize a human astrovirus strain apparently derived from naturally occurring recombination, detected in two different geographic locations and defined by EIA, RT-PCR, and phylogenetic methods.

The cDNA used to obtain sequence in the transition region of ORF1b and ORF2 was generated from a single RT-PCR generating a 1.2 kb product spanning the region and was sequenced for two different isolates (one Houston and one Mexican). High identity between these sequences, the obvious change in their
relationship to the reference HAstV-3 and HAstV-5 types in ORF1b and ORF2, and the lack of similar findings for 64 other HAstV strains [2] indicate that all four isolates are naturally occurring recombinants, probably derived from the same recombinant parent strain. Further, two different researchers characterized the Houston and Mexican isolates one year apart, which minimizes the chance that similarities were due to laboratory contamination.

HAstV co-infections with different antigenic types have been reported in epidemiological studies [11]. Co-infection by two different serotypes in one child provides an opportunity for recombination to occur. In addition, RNA viruses have a tendency toward recombination due to the nature of their Pol which naturally shift frame at the ORF1a/ORF1b junction [1, 6, 18, 19].

The preferred mechanism of recombination among +ssRNA viruses is a copy-choice mechanism where recombination occurs while the Pol enzyme switches from copying one RNA molecule (donor template) to another (acceptor template) without releasing the nascent strand [8]. RNA virus recombination events are classified as homologous, aberrant homologous, and non-homologous types depending on the location of the recombination site within the viral genome, and the precision of the event [10]. However, Nagy and Simon suggested a revision of the classification to sequence similarity-essential, similarity-assisted, and similarity-nonessential, based upon the mechanism of the generation and the nature of the recombinant product [18].

Studies with brome mosaic virus (BMV), a +ssRNA plant virus which is the most frequently used model for homologous RNA virus recombination, revealed several factors that are important in RNA recombination [14]. Those factors include the length of the identical region of the sequence, degree of sequence identity, the presence of AU-rich segments in the 3′ vs. relative GC-rich segments in the 5′ end, and stable hairpin structures [16, 17]. In the BMV model it was suggested that homologous recombination does not require similar secondary structures between the two recombining sites of the crossover. However, for the effective extension of the 3′-end, the Pol requires a hairpin structure in the acceptor RNA region [14, 19]. Koev et al. showed that the RNA secondary structure at the recombination site plays a unique role in the promoter recognition by the viral Pol in vivo in a recombination model of barley yellow dwarf virus [9].

Natural, homologous recombinants have been demonstrated among other +ssRNA viruses, including Picornaviridae [4] and Caliciviridae [7]. In the case of Picornaviridae, poliovirus with natural recombinant genome was isolated from vaccine-associated paralytic poliomyelitis in France, detected by double restriction fragment length polymorphism assay and confirmed by partial sequence analysis [4]. In the case of Caliciviridae, calicivirus with natural recombinant genome (Arg320) was identified from a diarrhea stool sample from Argentina and confirmed by sequence analysis [7]. In the case of Luteoviridae, Koev et al. proposed that recombination has occurred during luteovirus evolution, based on sequence analyses of genetic regions of the two major genera, Luteovirus and Polerovirus [9]. In Caliciviridae and Luteoviridae the recombination events were localized at the Pol/capsid junction, which is a transition between homologous
(non-structural) and divergent (structural) regions. Both of these viruses have subgenomic RNA encoding the capsid protein. In *Luteoviridae* it is proposed that recombination has occurred by *Pol* switching strand at the subgenomic promoter [9]. Neither natural nor artificially induced recombination has been detected among AstVs.

During the copy-choice mechanism of recombination the short AU rich sequence and stable hairpin structures of the donor template halt the synthesis on the nascent strand, thus providing an opportunity for the *Pol* to interact with the acceptor strand. These AU rich sequences also promote *Pol* slippage [15]. The stable hairpin structures act as replication enhancer in recruiting the *Pol* [17]. Long (or even short, depending on the virus) homologous sequences between the nascent RNA and the acceptor RNA facilitate the association of the nascent strand with the acceptor strand and function as a primer for the template switching and extension [14].

RNA sequence at the suggested recombination site we observed the following: high identity among all HAstV antigenic types (120nt 93–98% including a 52nt 99–100% region) and high identity with porcine AstV sequence (80–85%) in the 120 nt region. We did not locate definite AU rich segments in the 3′ vs. GC rich (or moderately AU-rich) segments in the 5′ end of this region. RNA secondary structure analysis predicted similar triple stable hairpin structures for all human and porcine strains. In summary, this proposed recombination site includes a long homologous region with stable hairpin structure, which supports the classification of our isolate as a homologous or similarity-assisted recombinant.

The recombination site in our case was in the junction of *Pol* and capsid region similar to other viruses (*Caliciviridae* and *Luteoviridae*) [7, 9]. Astroviruses transcribe a subgenomic RNA whose start point maps within the region of the suggested recombination junction, therefore we propose that polymerase recognition of the subgenomic RNA promoter may be the reason for the recombination event similar to *Luteoviridae*.

We conclude that phylogenetic analysis of different genomic regions of HAstV can provide contradictory genotyping unlike a previous observation [2]. In our case, the HAstV3-5 recombinant was widespread, apparently stable and pathogenic. Therefore molecular typing methods for AstVs should utilize sequence in the capsid region (ORF2) to correlate with antigenic typing methods.

RNA recombination may contribute to evolutionarily significant RNA rearrangements in astroviruses. Nucleic acid sequence homology and similar RNA secondary structure of HAstV and animal AstVs (porcine) in this transition region of ORF1b and ORF2 may provide a possibility of recombination among AstVs from different species should co-infections occur.

**Acknowledgements**

Supported in part by the National Institutes of Health (NICHHD13021 and NIAID AI45872). The authors thank Dr. Peter D. Nagy, Department of Plant Virology, University of Kentucky, Lexington, KY for his thoughtful comments on this manuscript.
References

1. Arnold JJ, Cameron CE (1999) Poliovirus RNA-dependent RNA polymerase (3Dpol) is sufficient for template switching in vitro. J Biol Chem 274: 2706–2716

2. Belliot G, Laveran H, Monroe SS (1997) Detection and genetic differentiation of human astroviruses: phylogenetic grouping varies by coding region. Arch Virol 142: 1323–1334

3. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). 3.5c ed. Department of Genetics, University of Washington, Seattle

4. Furione M, Guillot S, Otelea D, Balanant J, Candrea A, Crainic R (1993) Polioviruses with natural recombinant genomes isolated from vaccine-associated paralytic poliomyelitis. Virology 196: 199–208

5. Guerrero ML, Noel JS, Mitchell DK, Calva JJ, Morrow AL, Martinez J, Rosales G, Velazquez FR, Monroe SS, Glass RI, Pickering LK, Ruiz-Palacios GM (1998) A prospective study of astrovirus diarrhea of infancy in Mexico City. Pediatr Infect Dis J 17: 723–727

6. Jiang B, Monroe SS, Koonin EV, Stine SE, Glass RI (1993) RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proc Natl Acad Sci USA 90: 10539–10543

7. Jiang X, Espul C, Zhong WM, Cuello H, Matson DO (1999) Characterization of a novel human calicivirus that may be a naturally occurring recombinant. Arch Virol 144: 2377–2387

8. Kirkegaard K, Baltimore D (1986) The mechanism of RNA recombination in poliovirus. Cell 47: 433–443

9. Koev G, Mohan BR, Miller WA (1999) Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. J Virol 73: 2876–2885

10. Lai M (1996) Recombination in large RNA viruses: coronaviruses. Semin Virol: 381–388

11. Matsui M, Ushijima H, Hachiya M, Kakizawa J, Wen L, Oseto M, Morooka K, Kurtz JB (1998) Determination of serotypes of astroviruses by reverse transcription-polymerase chain reaction and homologies of the types by the sequencing of Japanese isolates. Microbiol Immunol 42: 539–547

12. Mitchell DK, Matson DO, Jiang X, Pickering LK (1995) Use of polymerase chain reaction for determination of astrovirus serotypes associated with diarrheal outbreaks among children attending day care centers (DCCs). Pediatr Res 37: 184A

13. Monroe SS, Jiang B, Stine SE, Koopmans M, Glass RI (1993) Subgenomic RNA sequence of human astrovirus supports classification of Astroviridae as a new family of RNA viruses. J Virol 67: 3611–3614

14. Nagy PD, Bujarski JJ (1995) Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. J Virol 69: 131–140

15. Nagy PD, Bujarski JJ (1997) Engineering of homologous recombination hotspots with AU-rich sequences in brome mosaic virus. J Virol 71: 3799–3810

16. Nagy PD, Ogiela C, Bujarski JJ (1999) Mapping sequences active in homologous RNA recombination in brome mosaic virus: prediction of recombination hot spots. Virology 254: 92–104

17. Nagy PD, Pogany J, Simon AE (1999) RNA elements required for RNA recombination function as replication enhancers in vitro and in vivo in a plus-strand RNA virus. Embo J 18: 5653–5665

18. Nagy PD, Simon AE (1997) New insights into the mechanisms of RNA recombination. Virology 235: 1–9
19. Nagy PD, Zhang C, Simon AE (1998) Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. Embo J 17: 2392–2403
20. Nicholas KB, Jr. NHB, Deerfield DWI (1997) GeneDoc: analysis and visualization of genetic variation. Embnewnews 4: 14
21. Noel JS, Lee TW, Kurtz JB, Glass RI, Monroe SS (1995) Typing of human astroviruses from clinical isolates by enzyme immunoassay and nucleotide sequencing. J Clin Microbiol 33: 797–801
22. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882
23. Walter JE, Berke T, Matson DO, Mitchell DK, Guerrero ML, Monroe SS, Pickering LK, Ruiz-Palacios GM (2001) Molecular epidemiology of human astrovirus diarrhea among children from a periurban community of Mexico City. J Infect Dis 183: 681–686
24. Willcocks MM, Brown TD, Madeley CR, Carter MJ (1994) The complete sequence of a human astrovirus. J Gen Virol 75: 1785–1788

Author’s address: Dr. D. K. Mitchell, Center for Pediatric Research, 855 West Brambleton Avenue, Norfolk, VA 23510-1001, U.S.A.; e-mail: dkmitchel@chkd.com

Received December 13, 2000