# Identification of a Novel Astrovirus (Astrovirus VA1) Associated with an Outbreak of Acute Gastroenteritis[▽]

Stacy R. Finkbeiner,[1]† Yan Li,[2]† Susan Ruone,[2] Christina Conrardy,[2] Nicole Gregoricus,[2]
Denise Toney,[3] Herbert W. Virgin,[1] Larry J. Anderson,[2] Jan Vinjé,[2]
David Wang,[1]* and Suxiang Tong[2]*

*Departments of Molecular Microbiology and Pathology and Immunology, Washington University School of Medicine, St. Louis,
Missouri 63110[1]; Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333[2]; and
Commonwealth of Virginia Division of Consolidated Laboratory Services, Richmond, Virginia 23219[3]*

**The etiology of a large proportion of gastrointestinal illness is unknown. In this study, random Sanger sequencing and pyrosequencing approaches were used to analyze fecal specimens from a gastroenteritis outbreak of unknown etiology in a child care center. Multiple sequences with limited identity to known astroviruses were identified. Assembly of the sequences and subsequent reverse transcription-PCR (RT-PCR) and rapid amplification of cDNA ends generated a complete genome of 6,586 nucleotides. Phylogenetic analysis demonstrated that this virus, named astrovirus VA1 (AstV-VA1), is highly divergent from all previously described astroviruses. Based on RT-PCR, specimens from multiple patients in this outbreak were unequivocally positive for Ast-VA1.**

Astroviruses consist of a family of small, single-stranded, positive-sense RNA viruses. Their genomes range from 6.1 to 7.3 kb in length (6, 14) and contain three open reading frames (ORFs) denoted ORF1a, -1b, and -2, which encode a serine protease, an RNA-dependent RNA polymerase (RdRP), and a capsid precursor protein, respectively (14). Astroviruses are known to infect a variety of species (3, 12, 20). In humans, eight serotypes have been described, which have been associated with up to ~10% of sporadic cases of diarrhea in children (2, 7, 10, 11, 17) and 0.5 to 15% of outbreaks (1, 13, 18). In addition, a highly divergent member of this family, astrovirus MLB1, was recently identified in patients with diarrhea (5, 6).

Significantly, the etiologies of 12 to 41% of all gastroenteritis outbreaks remain undetermined even after extensive testing, suggesting that there is a diagnostic gap (13, 18). In this paper, we applied mass sequencing to analyze specimens obtained from an unexplained outbreak of gastroenteritis at a child care center. We report the identification and complete genome sequencing of a novel astrovirus, referred to as astrovirus VA1 (AstV-VA1) (S. R. Finkbeiner, Y. Li, S. Ruone, C. Conrardy, N. Gregoricus, D. Toney, H. W. Virgin, L. J. Anderson, J. Vinjé, D. Wang, and S. Tong, U.S. patent application).

**Details of outbreak.** On 18 August 2008, the Eastern Shore Health District in Virginia was notified of cases of gastrointestinal illness at a child care center, with the outbreak lasting for a period of 2 to 3 weeks. Any attendee or staff member of the day care who had diarrhea and/or vomiting after 1 July 2008 fit the case definition for this outbreak. Control measures were put in place immediately at the center, including exclusion of symptomatic children, mandated testing of all symptomatic staff, testing of symptomatic children, and ultimately, temporary closing of the facility. By the conclusion of the outbreak, 26 patients fit the case definition. From these patients, fecal specimens from six patients (labeled A to F) (Table 1) were available for extensive testing. All six samples tested negative for known enteric parasites and enteric bacteria by standard microscopy analysis and culture. Similarly, all six samples tested negative for rotavirus (RotaClone enzyme immunoassay), norovirus, sapovirus, human astrovirus, and group F adenoviruses by reverse transcription-PCR (RT-PCR) (4, 16, 21), with the exception of samples B and F, which were intermittently positive at the limit of detection for human astrovirus.

**Genome amplification and sequencing.** Five of the fecal specimens (A to E) were analyzed independently in two laboratories by mass sequencing. At Washington University, total nucleic acid was extracted from diluted fecal specimens A, B, C, and D and randomly amplified as previously described (22), and the products were subjected to high-throughput pyrosequencing using the GS-FLX Titanium platform (Roche) (average of 12,730 reads per sample). We identified 313 unique high-quality sequence reads in sample B and 1,017 unique high-quality reads in sample C which were divergent from but most closely related to astroviruses, based on BLAST alignments. No astrovirus sequences were detected in sample A or D. A 6,376-nucleotide (nt) contig was assembled from the sequences detected in sample B, and four contigs totaling 6,026 nt were assembled from sample C. Because the overlapping sequences obtained in samples B and C were identical, the five original contigs were assembled to generate a 6,581-nt contig [excluding the poly(A) tail].

At CDC, total nucleic acid was extracted from samples A, B, C and E and randomly amplified as described previously (22).

TABLE 1. Epidemiologic data of six specimens from a child care center outbreak of acute gastroenteritis[a]

| Sample ID | Sex | Age | Date of (mo/day/yr): | | Symptom(s) | Result for AstV-VA1 with: | | | Result for others[c] |
|---|---|---|---|---|---|---|---|---|---|
| | | | Onset | Sample | | Sanger sequencing | Pyrosequencing | Real-time RT-PCR | |
| A | F | 43 yr | 8/19/08 | 8/19/08 | Diarrhea, vomiting, abdominal cramps | Neg | Neg | Weak Pos[b] | Neg |
| B | M | 10 mo | 8/18/08 | 8/18/08 | Diarrhea | Pos | Pos | Pos | Pos for human astrovirus |
| C | F | 10 mo | 8/04/08 | 8/19/08 | Diarrhea | Pos | Pos | Pos | Neg |
| D | M | 2 yr | 8/19/08 | 8/19/08 | Diarrhea, vomiting | NA | Neg | Weak Pos[b] | Neg |
| E | M | 2 yr | 8/19/08 | 8/19/08 | Diarrhea, vomiting | Neg | NA | Weak Pos[b] | Neg |
| F | M | 6 mo | 8/25/08 | 8/25/08 | Diarrhea | NA | NA | Pos | Pos for human astrovirus |

[a] ID, identification; Pos, positive; Neg, negative; F, female; M, male; NA, not applicable.
[b] Sample A showed positive results in late amplification cycles from one out of four separately extracted nucleic acid templates; samples D and E showed positive results in late amplification cycles from three out of four separately extracted nucleic acid templates.
[c] Other tests included those for the following: parasites; GI, GII, and GIV noroviruses; sapovirus and human astrovirus; rotavirus; and group F adenovirus.

Amplicons 300 to 800 bp in length were then cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA), and plasmids were sequenced using the Sanger method on an ABI Prism 3130 automated sequencer (Applied Biosystems, Foster City, CA). Three out of 96 clones from sample B and 69 out of 152 clones from sample C contained sequence signatures most closely related to previously known astroviruses by BLASTn similarity searches. Sequencing of 100 clones each from samples A and E yielded no clones with detectable similarity to astroviruses. The 69 clones from sample C were assembled into four contigs. Primers were then designed to generate a series of eight overlapping RT-PCR amplicons with an average size of ~900 bp that yielded a contig of 6,537 nt. In order to define the 5′ end of the genome, three independent rapid amplifications of 5′ cDNA ends were performed, and a total of 23 clones from these reactions were sequenced. All clones extended the genome by 49 nt and yielded the identical 5′ end sequence, suggesting that the genome was complete with a total length of 6,586 nt, excluding the poly(A) tail. The 3′ end was confirmed by rapid amplification of 3′ cDNA ends.

Comparison of the genome sequences generated by the two sequencing methods yielded nearly identical sequences, with the exception of five missing nucleotides at the 5′ end of the contig generated by pyrosequencing and three nucleotide substitution differences. These were resolved by direct PCR se-quencing to generate the final, corrected sequence, which was deposited in GenBank (accession number FJ973620).

**Genome analysis.** The genome of AstV-VA1 has three predicted ORFs as well as nontranslated regions (NTRs) at both the 5′ and 3′ ends of the genome. ORF1a and ORF2 were predicted by the NCBI ORF Finder (Table 2). The full coding region for ORF1b, which is produced by a $-1$ ribosomal frameshift during translation (8), was defined using the conserved heptameric "slippery sequence" (AAAAAAAC) near the end of ORF1a as the start site (8). The sequence AUUU GGAGNGGNGGACCNAAN$_{5-8}$AUGNC (start codon for ORF2 is italicized) located upstream of ORF2, which has been proposed as the promoter for subgenomic RNA synthesis in all previously known astroviruses (14), is also present in AstV-VA1 with only two nt differences. The 3′ NTR of nearly all astroviruses contains a highly conserved RNA secondary structure called the stem loop II-like motif (s2m) (9, 15). An alignment of the 150 nt just upstream of the poly(A) tail of AstV-VA1 with the 3′ NTR sequences of other astroviruses demonstrated that AstV-VA1 contained the highly conserved ~33-nt core of the s2m motif. The exact role of this motif is not understood; however, its presence in multiple viral families suggests it may play an important role in the astrovirus life cycle.

TABLE 2. Genome comparison of AstV-VA1 to other fully sequenced astroviruses

| Virus | Length of (nt): | | | | | |
|---|---|---|---|---|---|---|
| | Genome | 5′ NTR | ORF1a | ORF1b | ORF2 | 3′ NTR |
| Chicken astrovirus 1 | 6,927 | 15 | 3,017 | 1,533 | 2,052 | 305 |
| Turkey astrovirus 1 | 7,003 | 11 | 3,300 | 1,539 | 2,016 | 130 |
| Turkey astrovirus 2 | 7,325 | 21 | 3,378 | 1,584 | 2,175 | 196 |
| Mink astrovirus | 6,610 | 26 | 2,648 | 1,620 | 2,328 | 108 |
| Ovine astrovirus | 6,440 | 45 | 2,580 | 1,572 | 2,289 | 59 |
| Human astrovirus 1 | 6,813 | 85 | 2,763 | 1,560 | 2,361 | 80 |
| Human astrovirus 2 | 6,828 | 82 | 2,763 | 1,560 | 2,392 | 82 |
| Human astrovirus 4[a] | 6,723 | 84 | 2,763 | 1,548 | 2,316 | 81 |
| Human astrovirus 5[a] | 6,762 | 83 | 2,763 | 1,548 | 2,352 | 86 |
| Human astrovirus 8 | 6,759 | 80 | 2,766 | 1,557 | 2,349 | 85 |
| AstV-MLB1 | 6,171 | 14 | 2,364 | 1,536 | 2,271 | 58 |
| AstV-VA1 | 6,586 | 38 | 2,661 | 1,575 | 2,277 | 98 |

[a] Numbers were deduced from the full-length sequences.

A



100 amino acid
substitutions

B



100 amino acid
substitutions

C
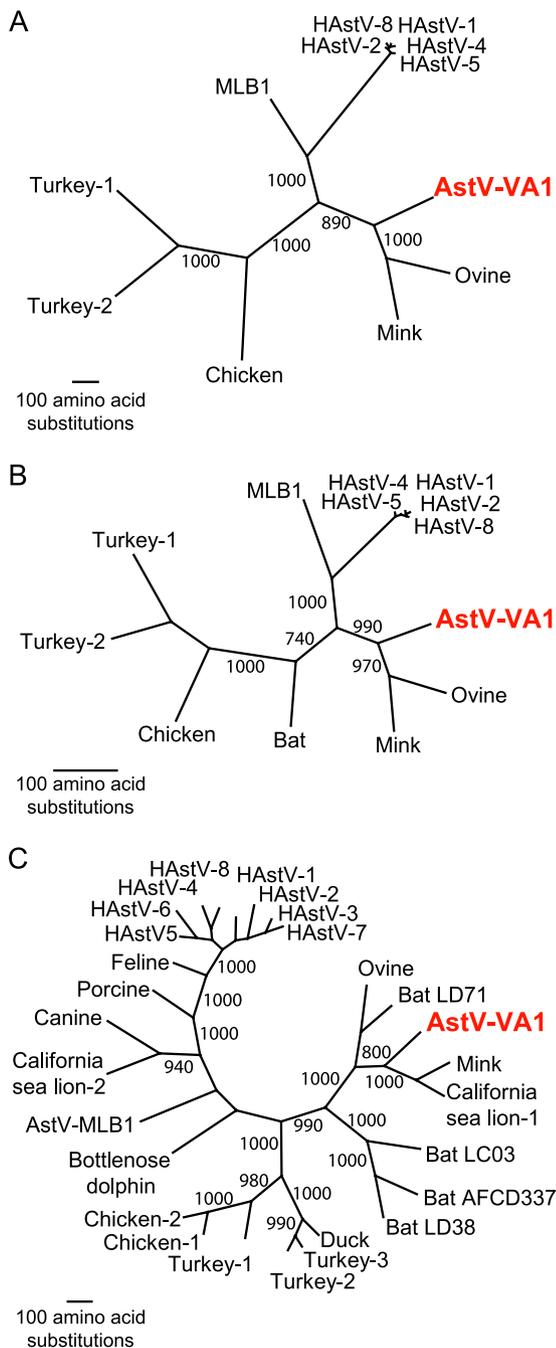


100 amino acid
substitutions

FIG. 1. Phylogenetic analysis of AstV-VA1 ORFs. Phylogenetic trees were generated in PAUP, using the maximum parsimony method with 1,000 bootstrap replicates. Significant bootstrap values are shown. (A) ORF1a serine protease. (B) ORF1b polymerase. (C) ORF2 capsid. HAstV, human astrovirus; Bat AFCD337, MpAstV/HK/AFCD337/06; Bat LD71, TmAstV/GX/LD71/07; Bat LC03, HpAstV/GX/LC03/07; Bat LD38, TmAstV/GX/LD38/07.

**Phylogenetic analysis.** ClustalX (version 1.83) was used to align each complete ORF of Ast-VA1 with the respective available complete ORFs of other astroviruses in GenBank. Maximum parsimony trees were then generated using PAUP with 1,000 bootstrap replicates (19). This analysis demonstrated that AstV-VA1 was highly divergent from but most closely

related to mink and ovine astroviruses in ORF1a and ORF1b (Fig. 1A and B). In the capsid region, for which more astrovirus sequences are available, AstV-VA1 was most similar to mink and California sea lion astroviruses (Fig. 1C).

In terms of sequence identity, as expected, ORF1b was the most highly conserved region, sharing 61% amino acid identity to mink astrovirus and 62% to ovine astrovirus. The ORF1a (serine protease) coding region was more divergent, with 39% and 40% amino acid identities with ovine astrovirus and mink astrovirus, respectively. In ORF2, AstV-VA1 shared 41% amino acid identity to mink astrovirus and 41% to California sea lion astrovirus 1.

**RT-PCR screening for AstV-VA1.** Real-time RT-PCR and semi-nested RT-PCR assays were developed, targeting regions in ORF1b and ORF2 of AstV-VA1, respectively. All six samples were tested with both assays (Table 1). Four independent nucleic acid extractions of each sample were prepared. Each extraction of samples B, C, and F was unequivocally positive in both assays, with threshold cycle ($C_t$) values in the real-time RT-PCR assay ranging from 18 to 20, suggesting that a high copy number of Ast-VA1 was present in those samples. The other three samples were intermittently weakly positive in the semi-nested RT-PCR assay (A, 1/4 extractions; D, 3/4 extractions; E, 1/3 extractions) and in the real-time RT-PCR assay (A, 1/4 extractions; D and E, 3/4 extractions). For the real-time RT-PCR assay, in the instances where these three samples were positive, the $C_t$ values were near the limit of detection, ranging from 34 to 42. These results suggest that samples A, D, and E may contain very low copy numbers of AstV-VA1 RNA, which may explain the variation in results for the four independent nucleic acid extractions. Negative controls included on each run were all negative. The 250-bp amplicon generated by the semi-nested PCR assay was confirmed as AstV-VA1 in all samples by sequencing.

Despite the availability of improved molecular diagnostic methods for an increasing panel of gastroenteritis agents in humans, the etiology of 12 to 41% of the outbreaks of gastroenteritis remains unexplained (13, 18). In this study, we identified a novel astrovirus (AstV-VA1) in fecal samples from an outbreak of acute gastroenteritis. Complete genome sequencing and phylogenetic analysis demonstrated that AstV-VA1 was highly divergent from all previously described astroviruses, including the eight human astrovirus serotypes and the recently described astrovirus MLB1 (AstV-MLB1) (6). The discovery of AstV-VA1 following the recent identification of AstV-MLB1 clearly demonstrates that a much greater diversity of astroviruses exists in humans than was previously recognized.

The detection of AstV-VA1 at high copy numbers in three out of six samples (and potentially at very low levels in the other three samples) from this outbreak suggests a potential association between AstV-VA1 and diarrheal illness. However, because of the limited number of samples available for analysis in this cluster, further studies defining the frequency of detection of AstV-VA1 in samples from individuals with and without acute gastroenteritis are needed to define the role of AstV-VA1 in human diarrheal disease.

**Nucleotide sequence accession number.** The nucleotide sequence determined in this study was deposited in the GenBank database under accession number FJ973620.

## REFERENCES

1. **Akihara, S., T. G. Phan, T. A. Nguyen, G. Hansman, S. Okitsu, and H. Ushijima.** 2005. Existence of multiple outbreaks of viral gastroenteritis among infants in a day care center in Japan. Arch. Virol. **150:**2061–2075.

2. **Caracciolo, S., C. Minini, D. Colombrita, I. Foresti, M. Avolio, G. Tosti, S. Fiorentini, and A. Caruso.** 2007. Detection of sporadic cases of norovirus infection in hospitalized children in Italy. New Microbiol. **30:**49–52.

3. **Chu, D. K., L. L. Poon, Y. Guan, and J. S. Peiris.** 2008. Novel astroviruses in insectivorous bats. J. Virol. **82:**9107–9114.

4. **Davidson, G., R. Townley, R. F. Bishop, I. Holmes, and B. Ruck.** 1975. Importance of a new virus in acute sporadic enteritis in children. Lancet **i:**242–246.

5. **Finkbeiner, S. R., A. F. Allred, P. I. Tarr, E. J. Klein, C. D. Kirkwood, and D. Wang.** 2008. Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog. **4:**e1000011.

6. **Finkbeiner, S. R., C. D. Kirkwood, and D. Wang.** 2008. Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. Virol. J. **5:**117.

7. **Glass, R. I., J. Noel, D. Mitchell, J. E. Herrmann, N. R. Blacklow, L. K. Pickering, P. Dennehy, G. Ruiz-Palacios, M. L. de Guerrero, and S. S. Monroe.** 1996. The changing epidemiology of astrovirus-associated gastroenteritis: a review. Arch. Virol. Suppl. **12:**287–300.

8. **Jiang, B., S. S. Monroe, E. V. Koonin, S. E. Stine, and R. I. Glass.** 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.

9. **Jonassen, C. M., T. O. Jonassen, and B. Grinde.** 1998. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J. Gen. Virol. **79**(Pt 4)**:**715–718.

10. **Kirkwood, C. D., R. Clark, N. Bogdanovic-Sakran, and R. F. Bishop.** 2005. A 5-year study of the prevalence and genetic diversity of human caliciviruses associated with sporadic cases of acute gastroenteritis in young children admitted to hospital in Melbourne, Australia (1998–2002). J. Med. Virol. **77:**96–101.

11. **Klein, E. J., D. R. Boster, J. R. Stapp, J. G. Wells, X. Qin, C. R. Clausen, D. L. Swerdlow, C. R. Braden, and P. I. Tarr.** 2006. Diarrhea etiology in a children's hospital emergency department: a prospective cohort study. Clin. Infect. Dis. **43:**807–813.

12. **Koci, M. D., and S. Schultz-Cherry.** 2002. Avian astroviruses. Avian Pathol. **31:**213–227.

13. **Lyman, W. H., J. F. Walsh, J. B. Kotch, D. J. Weber, E. Gunn, and J. Vinje.** 2009. Prospective study of etiologic agents of acute gastroenteritis outbreaks in child care centers. J. Pediatr. **154:**253–257.

14. **Mendez, E., and C. F. Arias.** 2007. Astroviruses, p. 981–1000. *In* D. M. Knipe and P. M. Howley (ed.), Fields virology, 5th ed., vol. 1. Lippincott Willliams & Wilkins, Philadelphia, PA.

15. **Monceyron, C., B. Grinde, and T. O. Jonassen.** 1997. Molecular characterisation of the 3′-end of the astrovirus genome. Arch. Virol. **142:**699–706.

16. **Oka, T., K. Katayama, G. S. Hansman, T. Kageyama, S. Ogawa, F. T. Wu, P. A. White, and N. Takeda.** 2006. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. J. Med. Virol. **78:**1347–1353.

17. **Soares, C. C., M. C. Maciel de Albuquerque, A. G. Maranhao, L. N. Rocha, M. L. Ramirez, F. J. Benati, C. Timenetsky Mdo, and N. Santos.** 2008. Astrovirus detection in sporadic cases of diarrhea among hospitalized and non-hospitalized children in Rio De Janeiro, Brazil, from 1998 to 2004. J. Med. Virol. **80:**113–117.

18. **Svraka, S., E. Duizer, H. Vennema, E. de Bruin, B. van der Veer, B. Dorresteijn, and M. Koopmans.** 2007. Etiological role of viruses in outbreaks of acute gastroenteritis in The Netherlands from 1994 through 2005. J. Clin. Microbiol. **45:**1389–1394.

19. **Swofford, D. L.** 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.

20. **Toffan, A., C. M. Jonassen, C. De Battisti, E. Schiavon, T. Kofstad, I. Capua, and G. Cattoli.** 4 May 2009, posting date. Genetic characterization of a new astrovirus detected in dogs suffering from diarrhea. Vet. Microbiol., doi: 10.1016/j.vetmic.2009.04.031.

21. **Trujillo, A. A., K. A. McCaustland, D. P. Zheng, L. A. Hadley, G. Vaughn, S. M. Adams, T. Ando, R. I. Glass, and S. S. Monroe.** 2006. Use of TaqMan real-time reverse transcription-PCR for rapid detection, quantification, and typing of norovirus. J. Clin. Microbiol. **44:**1405–1412.

22. **Wang, D., A. Urisman, Y. T. Liu, M. Springer, T. G. Ksiazek, D. D. Erdman, E. R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J. P. Latreille, R. K. Wilson, D. Ganem, and J. L. DeRisi.** 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. **1:**E2.