



Complete Genome Sequences of Human Astrovirus Prototype Strains (Types 1 to 8)

 Christina J. Castro,^{a,b} Emily Reynolds,^{a,b} Stephan S. Monroe,^c Rachel L. Marine,^b Jan Vinjé^b

^aOak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

^bDivision of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

^cOffice of Laboratory Science and Safety, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT We report the complete genome sequences of the eight human astrovirus Oxford prototype strains. These sequences share 94.9% to 99.9% nucleotide identity with open reading frame 2 (ORF2) genes of astrovirus genomes previously deposited in GenBank and include the first complete genome of human astrovirus type 7.

Astroviruses belong to the family *Astroviridae*, which is composed of two genera, *Mamastrovirus* and *Avastrovirus*, which include viruses that infect mammals and birds, respectively. They are small (≈ 28 to 35 nm) star-shaped nonenveloped icosahedral viruses with a nonsegmented single-stranded positive-sense RNA genome (1). The genome is 6,700 to 7,000 nucleotides long and contains three open reading frames (ORFs). ORF1a and ORF1b encode the nonstructural protease and RNA-dependent RNA polymerase proteins, respectively, while ORF2 encodes the viral structural protein (1), which is commonly used for genotyping of astroviruses.

Clinical symptoms of human astrovirus (HAstV) infections include headache, diarrhea, vomiting, and abdominal pain. While infections usually resolve without specific treatment, severe dehydration leading to hospitalization has been described (2–5). HAstVs were first reported in 1975 (2) after electron microscopy (EM) analysis of stool samples from hospitalized infants with acute gastroenteritis. HAstV serotypes were initially differentiated by immune EM (6, 7), and in the early 1990s, based on sequences of the capsid gene, they were classified in the family *Astroviridae* (8). To date, eight serotypes have been described (6, 7, 9). HAstV accounts for 2% to 9% of all sporadic cases of acute gastroenteritis in children, and HAstV-1 has been reported as the most common strain globally (1, 10–12). Coinfection with other viruses, such as rotavirus, norovirus, or sapovirus, has been reported frequently (13).

Fewer than 30 complete HAstV genome sequences are available in GenBank, including types 1 to 6 and 8. We report the complete genomes of the original HAstV Oxford reference strains, types 1 to 8.

The first HAstVs that were successfully cultured in the laboratory of John Kurtz (Oxford, United Kingdom) have since been referred to as the Oxford reference strains. The eight isolates were originally obtained from John Kurtz and John Herrmann (University of Massachusetts Medical School, Worcester, MA) and were cultured in LLCMK2D cells as described previously (14). Astrovirus RNA was extracted using the QIAamp viral RNA minikit followed by on-column DNase digestion (Qiagen). Sequence-independent, single-primer amplification (15, 16) was used to produce amplicons that were processed using the Nextera XT library preparation kit. The resulting libraries were sequenced on an Illumina MiSeq 500-cycle paired-end run.

A custom in-house bioinformatics pipeline (17) was used to process raw FASTQ data and to *de novo* assemble each isolate. Within the pipeline, the preprocessing steps for

Citation Castro CJ, Reynolds E, Monroe SS, Marine RL, Vinjé J. 2019. Complete genome sequences of human astrovirus prototype strains (types 1 to 8). *Microbiol Resour Announc* 8:e01611-18. <https://doi.org/10.1128/MRA.01611-18>.

Editor Jelle Matthijnsens, KU Leuven

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Jan Vinjé, jvinje@cdc.gov.

Received 3 December 2018

Accepted 21 January 2019

Published 14 February 2019

the FASTQ raw reads were as follows: host removal using Bowtie 2 v2.3.3.1 (18–20) followed by primer and adapter trimming and Phred quality score filtering (removing those with a score of <20) using Cutadapt v1.8.3 (21) and, finally, removal of duplicate reads using the Python script Dedup.py (22). The remaining FASTQ reads were assembled into contigs using the *de novo* assembler SPAdes v3.7.0 (23) and contigs were then classified into taxonomic groups using NCBI BLAST+ v2.6.0 (24). Consensus genome sequences were verified through read mapping and annotated using Geneious vR11. Complete genome sequences with an average read coverage per genome ranging from 600 × to 2,300 × were generated. The total postprocessing FASTQ reads per sample ranged from 33,000 to 112,000, with the percentage of target viral reads between 88.7 and 96.9%.

Compared to the closest corresponding complete HAstV genomes available from GenBank, pairwise nucleotide identities from the Oxford strains ranged from 99.8% for HAstV-1 (GenBank accession number [L23513](#)), 99.6% for HAstV-2 (GenBank accession number [L13745](#)), 97.6% for HAstV-3 (GenBank accession number [AF141381](#)), 94.6% for HAstV-4 (GenBank accession number [AY720891](#)), 93.7% for HAstV-5 (GenBank accession number [JQ403108](#)), 95.2% for HAstV-6 (GenBank accession number [HM237363](#)), and 97.3% for HAstV-8 (GenBank accession number [AF260508](#)). The ORF2 of HAstV-7 had a 99.9% pairwise nucleotide identity with a partial ORF2 HAstV-7 sequence available from GenBank (accession number [Y08632](#)).

Data availability. The HAstV Oxford reference genome sequences (types 1 to 8) have been deposited in GenBank with the accession numbers [MK059949](#) to [MK059956](#). The postprocessed FASTQ reads have been deposited in the Sequence Read Archive with the run accession numbers [SRR8444451](#) to [SRR8444458](#).

ACKNOWLEDGMENTS

This research was supported in part by an appointment to the Research Participation Program at the CDC administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy and the CDC.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

- Bosch A, Pintó RM, Guix S. 2014. Human astroviruses. *Clin Microbiol Rev* 27:1048–1074. <https://doi.org/10.1128/CMR.00013-14>.
- Madeley CR, Cosgrove BP. 1975. Letter: 28 nm particles in faeces in infantile gastroenteritis. *Lancet* 2:451–452. [https://doi.org/10.1016/S0140-6736\(75\)90858-2](https://doi.org/10.1016/S0140-6736(75)90858-2).
- Siqueira JAM, Oliveira DS, de Carvalho TCN, Portal TM, Justino MCA, da Silva LD, Resque HR, Gabbay YB. 2017. Astrovirus infection in hospitalized children: molecular, clinical and epidemiological features. *J Clin Virology* 94:79–85. <https://doi.org/10.1016/j.jcv.2017.07.014>.
- Zaraket H, Abou-El-Hassan H, Kreidieh K, Soudani N, Ali Z, Hammadi M, Reslan L, Ghanem S, Hajar F, Inati A, Rajab M, Fakhouri H, Ghanem B, Baasiri G, Melhem NM, Dbaibo G. 2017. Characterization of astrovirus-associated gastroenteritis in hospitalized children under five years of age. *Infect Genet Evol* 53:94–99. <https://doi.org/10.1016/j.meegid.2017.05.016>.
- Daniel-Wayman S, Fahle G, Palmore T, Green KY, Prevots DR. 2018. Norovirus, astrovirus, and sapovirus among immunocompromised patients at a tertiary care research hospital. *Diagn Microbiol Infect Dis* 92:143–146. <https://doi.org/10.1016/j.diagmicrobio.2018.05.017>.
- Lee TW, Kurtz JB. 1982. Human astrovirus serotypes. *J Hyg (Lond)* 89:539–540. <https://doi.org/10.1017/S0022172400071114>.
- Kurtz JB, Lee TW. 1984. Human astrovirus serotypes. *Lancet* 324:1405. [https://doi.org/10.1016/S0140-6736\(84\)92101-9](https://doi.org/10.1016/S0140-6736(84)92101-9).
- 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.
- Lee TW, Kurtz JB. 1994. Prevalence of human astrovirus serotypes in the Oxford region 1976–92, with evidence for two new serotypes. *Epidemiol Infect* 112:187–193. <https://doi.org/10.1017/S0950268800057551>.
- Guix S, Caballero S, Villena C, Bartolomé R, Latorre C, Rabella N, Simó M, Bosch A, Pintó RM. 2002. Molecular epidemiology of astrovirus infection in Barcelona, Spain. *J Clin Microbiol* 40:133–139. <https://doi.org/10.1128/JCM.40.1.133-139.2002>.
- Nguyen TA, Hoang L, Pham LD, Hoang KT, Mizuguchi M, Okitsu S, Ushijima H. 2008. Identification of human astrovirus infections among children with acute gastroenteritis in the southern part of Vietnam during 2005–2006. *J Med Virol* 80:298–305. <https://doi.org/10.1002/jmv.21032>.
- Palombo EA, Bishop RF. 1996. Annual incidence, serotype distribution, and genetic diversity of human astrovirus isolates from hospitalized children in Melbourne, Australia. *J Clin Microbiol* 34:1750–1753.
- Jeong HS, Jeong A, Cheon D-S. 2012. Epidemiology of astrovirus infection in children. *Korean J Pediatr* 55:77–82. <https://doi.org/10.3345/kjp.2012.55.3.77>.
- Willcocks MM, Carter MJ, Laidler FR, Madeley CR. 1990. Growth and characterisation of human faecal astrovirus in a continuous cell line. *Arch Virol* 113:73–81. <https://doi.org/10.1007/BF01318354>.
- Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, Spiro DJ. 2008. Viral genome sequencing by random priming methods. *BMC Genomics* 9:5. <https://doi.org/10.1186/1471-2164-9-5>.
- Reyes GR, Kim JP. 1991. Sequence-independent, single-primer amplifi-

- cation (SISPA) of complex DNA populations. *Mol Cell Probes* 5:473–481. [https://doi.org/10.1016/S0890-8508\(05\)80020-9](https://doi.org/10.1016/S0890-8508(05)80020-9).
17. Montmayeur AM, Ng TF, Schmidt A, Zhao K, Magana L, Iber J, Castro CJ, Chen Q, Henderson E, Ramos E, Shaw J, Tatusov RL, Dybdahl-Sissoko N, Endegue-Zanga MC, Adeniji JA, Oberste MS, Burns CC. 2017. High-throughput next-generation sequencing of polioviruses. *J Clin Microbiol* 55:606–615. <https://doi.org/10.1128/JCM.02121-16>.
 18. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
 19. Langmead B, Wilks C, Antonescu V, Charles R. 2018. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty648>.
 20. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357. <https://doi.org/10.1038/nmeth.1923>.
 21. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
 22. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, Delwart EL. 2015. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* 43:e46. <https://doi.org/10.1093/nar/gkv002>.
 23. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.