



Near-Complete Genome Sequences of Eight Human Astroviruses Recovered from Diarrheal Stool Samples of Hospitalized Children in Coastal Kenya in 2019

 Arnold W. Lambisia,^a My V. T. Phan,^b Zaydah R. de Laurent,^a Matthew Cotten,^b D. James Nokes,^{a,c}  Charles N. Agoti^{a,d}

^aKenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

^bUK Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit, Entebbe, Uganda

^cSchool of Life Sciences and Zeeman Institute (SBIDER), University of Warwick, Coventry, United Kingdom

^dSchool of Health and Human Sciences, Pwani University, Kilifi, Kenya

ABSTRACT Here, using a sequence-independent sequencing approach (M. V. Phan, P. Hong Anh, N. Van Cuong, B. Oude Munnink, et al., *Virus Evol* 2:vew027, 2016, <https://doi.org/10.1093/ve/vew027>), we determined human astrovirus (HAstV) genome sequences from eight diarrheal stool samples collected in coastal Kenya in 2019. Phylogenetic analysis identified the following 4 genotypes: HAstV-1 ($n = 4$), HAstV-2 ($n = 1$), HAstV-3 ($n = 1$), and HAstV-5 ($n = 2$).

Human astroviruses (HAstVs) (family *Astroviridae*) are nonenveloped, 7-kb positive-sense, single-stranded RNA genome viruses (1) and are among the top 5 viral causes of childhood diarrhea globally (2). HAstV clinical isolates are classified into classic HAstVs (HAstV-1 to HAstV-8), HAstV-MLB, and HAstV-VA/HMO (1).

In Kenya and other African settings, HAstV positivity in children with diarrhea as one of their illness symptoms ranges from 2.7% to 10.3% (3–5). To date, there are no complete or near-complete ($\geq 90\%$ genome coverage) HAstV genome sequences from East Africa in the GenBank database (6). Analysis of HAstV genome sequences may facilitate optimization of molecular diagnostics and tracking the spread of HAstVs (7). Here, we utilized sequence-independent single-primer amplification (SISPA) sequencing to generate new HAstV genome sequences from positive reverse transcription-quantitative PCR (RT-PCR) (5) samples collected from children hospitalized with diarrhea in Kilifi, Kenya.

Total nucleic acid (TNA) was extracted from the 10 stool specimens using the QIAamp fast DNA stool minikit (Qiagen, Manchester, United Kingdom). The TNA was treated with Turbo DNase (Invitrogen, Carlsbad, CA), and first-strand synthesis was performed with FR26RV-ENDOH primers (8). Second-strand DNA synthesis was performed with Klenow fragment 3' to 5' exo- (New England BioLabs). To achieve a nonselective nucleic acid amplification, double-stranded DNA (dsDNA) was primed with the FR20RV primer (5'-GCCGGAGCTCTGCAGATATC-3'), complementary to the FR26RV-ENDOH primers at the 5' end (9), and amplified using SuperScript III with the Platinum *Taq* DNA polymerase kit (Qiagen) as per the manufacturer's protocol. The PCR product was used to prepare Illumina barcoded libraries using the Illumina DNA Flex kit and sequenced in one run using the Illumina MiSeq machine generating 75-bp paired-end reads. Sequencing adapters and low-quality bases (Phred score, < 30) were trimmed/removed from the short-read data using QUASR v.7.03 (10). Reference HAstV-1, HAstV-2, HAstV-3, and HAstV-5 genome sequences (GenBank accession numbers [JF327666](https://doi.org/10.1093/ve/vew027), [KF039911](https://doi.org/10.1093/ve/vew027), [MN444721](https://doi.org/10.1093/ve/vew027), and [MF684776](https://doi.org/10.1093/ve/vew027), respectively) were used for reference-guided assembly and to transfer annotations to the assembled genomes using the inbuilt Geneious mapper

Citation Lambisia AW, Phan MVT, de Laurent ZR, Cotten M, Nokes DJ, Agoti CN. 2021. Near-complete genome sequences of eight human astroviruses recovered from diarrheal stool samples of hospitalized children in coastal Kenya in 2019. *Microbiol Resour Announc* 10:e00162-21. <https://doi.org/10.1128/MRA.00162-21>.

Editor Jelle Matthijnssens, KU Leuven

Copyright © 2021 Lambisia et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Arnold W. Lambisia, alambisia@kemri-wellcome.org.

Received 11 February 2021

Accepted 16 March 2021

Published 15 April 2021

TABLE 1 Characteristics of human astrovirus genomes from coastal Kenya in 2019

Strain	Type	Collection date (day-mo-yr)	C_T value ^a	Age (mo)	Sex	Symptom(s) ^b	Genome length (nt) ^c	Total no. of raw reads	No. of mapped reads	Avg depth ^d	Genome coverage (%) ^e	GC content (%)	Pairwise identity to reference (%)	GenBank accession no.	Reference genome length (nt)
KLF/ASV/001	HAsTV1	13/4/2019	19.5	25	Female	D + V	6,115	2,014,832	179,076	138	90.24	44.3	97.2	MW485038	6,776
KLF/ASV/008	HAsTV1	26/4/2019	22.8	23	Male	D + V	6,776	1,297,222	5,673	53	100.00	44.9	97.4	MW485040	6,776
KLF/ASV/010	HAsTV1	18/7/2019	22.5	15	Male	D + V	6,398	1,317,294	878	9	94.42	47.9	97.2	MW485041	6,776
KLF/ASV/006	HAsTV1	23/7/2019	21.4	8	Male	D + V	6,698	3,015,006	3,744	39	98.85	45.0	97.3	MW485039	6,776
KLF/ASV/009	HAsTV1	10/6/2019	24.0	27	Female	D + V	5,342	1,912,848	388	5	78.84				6,776
KLF/ASV/004	HAsTV1	19/6/2019	22.2	10	Female	D + V	4,788	1,952,244	195	3	70.66				6,776
KLF/ASV/005	HAsTV2	19/6/2019	23.9	12	Male	D + V	6,725	1,581,106	13,121	158	99.22	44.2	90.3	MW485042	6,778
KLF/ASV/007	HAsTV3	15/4/2019	26.2	22	Female	D	6,747	2,302,640	10,137	100	99.37	44.0	94.1	MW485043	6,790
KLF/ASV/002	HAsTV5	1/6/2019	24.3	9	Female	D	6,666	2,906,060	1,769	18	97.99	43.6	98.3	MW485044	6,803
KLF/ASV/003	HAsTV5	1/6/2019	22.4	24	Male	D + V	6,361	3,046,380	1,188	13	93.50	43.7	98.5	MW485045	6,803

^aThe real-time RT-PCR (rRT-PCR) assay, including primers and probe sequences used for HAsTV detection, has been described previously (6). C_T cycle threshold.

^bObjective evidence of a diarrheal disease. D, diarrhea; V, vomiting.

^cnt, nucleotide.

^dCalculated by dividing the per-position coverage output by respective genome length.

^eCalculated by dividing the genome length by the respective reference genome length.

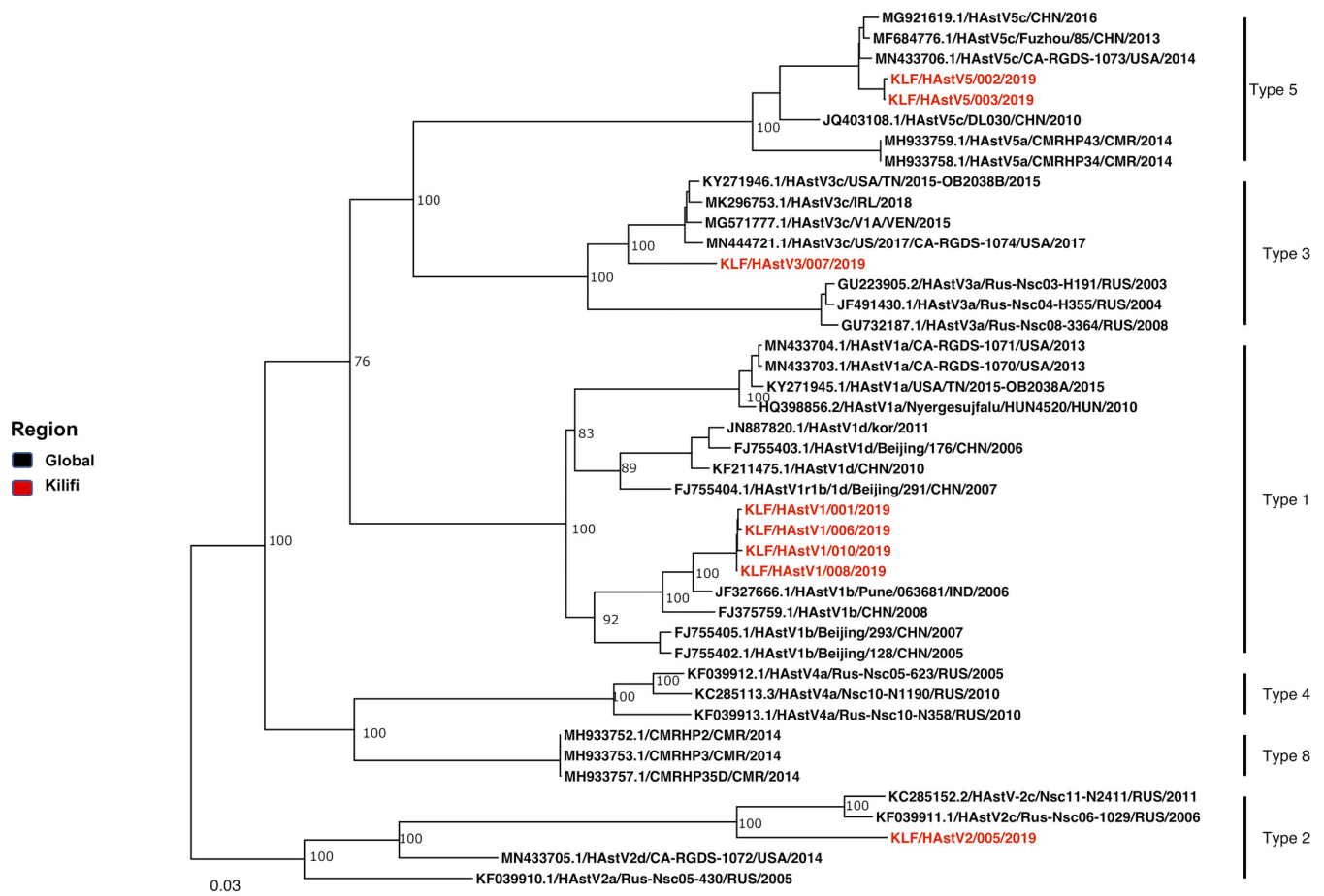


FIG 1 Maximum likelihood phylogenetic tree based on the open reading frame (ORF) sequences of the eight classical HAstVs (>90% genome coverage) identified in this study and representative strains from GenBank. The tree was constructed using IQ-Tree v.2.0.6 (13) with standard model selection. Bar indicates nucleotide substitutions per site. Red and black show HAstVs identified in this study and globally, respectively.

and annotation tools, respectively, on Geneious Prime v.2019.2.3 (11). MAFFT v.7.3.13 (12) was used for nucleotide coding sequence alignment, and maximum likelihood phylogenies were reconstituted in IQ-Tree v.2.0.6 (13) with standard model selection. Written informed consent for study participation was obtained from parents/guardians of the enrolled children, and the study protocol was approved by the KEMRI Scientific and Ethics Review Unit (SSC 2861 and SERU CGMRC/113/3624).

Patient demographics and sequencing output characteristics for the 10 samples are provided in Table 1. Eight samples yielded a consensus sequence covering >90% of the HAstV full-length genome. A maximum likelihood phylogeny of these eight near-complete genomes, including all publicly available HAstV genomes, is shown in Fig. 1. The new Kilifi sequences clustered with four different types of classical HAstVs, namely, HAstV-1 ($n = 4$), HAstV-2 ($n = 1$), HAstV-3 ($n = 1$), and HAstV-5 ($n = 2$). Both the HAstV-1 ($n = 4$) and HAstV-5 ($n = 2$) genomes had > 99% nucleotide similarity within their respective types. These new near-complete HAstV genomes from coastal Kenya increase available HAstV genomic data to support future molecular studies and local diagnostic methods.

Data availability. The raw sequence data were deposited in the Sequence Read Archive (SRA) under BioProject accession number [PRJNA692787](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA692787) and BioSample accession numbers [SAMN17370496](https://www.ncbi.nlm.nih.gov/biosample/SAMN17370496) to [SAMN17370503](https://www.ncbi.nlm.nih.gov/biosample/SAMN17370503). The genome sequences generated here were deposited in GenBank under accession numbers [MW485038](https://www.ncbi.nlm.nih.gov/nuccore/MW485038) to [MW485045](https://www.ncbi.nlm.nih.gov/nuccore/MW485045).

ACKNOWLEDGMENTS

We thank the study participants who provided the material we analyzed here.

This study was funded by The Wellcome Trust (102975 and 203077) and the Initiative to Develop African Research Leaders (IDeAL) through the DELTAS Africa Initiative (DEL-15-003). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from The Wellcome Trust (107769/Z/10/Z) and the UK government.

The views expressed in this publication are ours and not necessarily those of AAS, NEPAD Agency, The Wellcome Trust, or the UK government.

This paper is published with the permission of the director of KEMRI.

REFERENCES

- Donato C, Vijaykrishna D. 2017. The broad host range and genetic diversity of mammalian and avian astroviruses. *Viruses* 9:102. <https://doi.org/10.3390/v9050102>.
- Olorategui MP, Rouhani S, Yori PP, Salas MS, Trigos DR, Mondal D, Bodhidatta L, Platts-Mills J, Samie A, Kabir F, Lima A, Babji S, Shrestha SK, Mason CJ, Kalam A, Bessong P, Ahmed T, Mduma E, Bhutta ZA, Lima I, Ramdass R, Moulton LH, Lang D, George A, Zaidi AKM, Kang G, Houpt ER, Kosek MN, on behalf of the MAL-ED Network. 2018. Astrovirus infection and diarrhea in 8 countries. *Pediatrics* 141:e20171326. <https://doi.org/10.1542/peds.2017-1326>.
- Lekana-Douki SE, Kombila-Koumavor C, Nkoghe D, Drosten C, Drexler JF, Leroy EM. 2015. Molecular epidemiology of enteric viruses and genotyping of rotavirus A, adenovirus and astrovirus among children under 5 years old in Gabon. *Int J Infect Dis* 34:90–95. <https://doi.org/10.1016/j.ijid.2015.03.009>.
- Nguekeng Tsague B, Mikounou Louya V, Ntoumi F, Adedjoja A, Vouvougui CJ, Peko SM, Abena AA. 2020. Occurrence of human astrovirus associated with gastroenteritis among Congolese children in Brazzaville, Republic of Congo. *Int J Infect Dis* 95:142–147. <https://doi.org/10.1016/j.ijid.2020.02.056>.
- Lambisia AW, Onchaga S, Murunga N, Lewa CS, Nyanjom SG, Agoti CN. 2020. Epidemiological trends of five common diarrhea-associated enteric viruses pre-and post-rotavirus vaccine introduction in coastal Kenya. *Pathogens* 9:660. <https://doi.org/10.3390/pathogens9080660>.
- Federhen S. 2012. The NCBI taxonomy database. *Nucleic Acids Res* 40: D136–D143. <https://doi.org/10.1093/nar/gkr1178>.
- Yinda CK, Shi C, Deboutte W, Conceição-Neto N, Van Ranst M, Beller L, Maes P, Vanhulle E, Ghogomu SM, Matthijssens J. 2019. Gut virome analysis of Cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *mSphere* 4:e00585-18. <https://doi.org/10.1128/mSphere.00585-18>.
- Nguyen AT, Tran TT, Hoang VMT, Nghiem NM, Le NNT, Le TTM, Phan QT, Truong KH, Le NNT, Ho VL, Do VC, Ha TM, Nguyen HT, Nguyen CVV, Thwaites G, Van Doorn HR, Van Le T. 2016. Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. *Virol J* 13:125. <https://doi.org/10.1186/s12985-016-0580-9>.
- 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>.
- Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P. 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc B Biol Sci* 368:20120205. <https://doi.org/10.1098/rstb.2012.0205>.
- Biomatters Limited. 2018. Geneious—molecular biology and NGS analysis tools. Biomatters Limited, Auckland, New Zealand.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.