



Near-Complete Human Sapovirus Genome Sequences from Kenya

Marta Diez-Valcarce,^a Anna Montmayeur,^b Roman Tatusov,^c Jan Vinjé^d

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

^bSynergy, Atlanta, Georgia, USA

^cIHRC, Atlanta, Georgia, USA

^dDivision of Viral Diseases, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT We report five near-complete sapovirus genome sequences, including GI.3, GII.2, and GII.6 and two novel GII.NA (not assigned) strains. These new sequences expand the collection of human sapoviruses, allowing for a more accurate phylogenetic analysis of circulating strains and for designing broadly reactive primers for their detection and typing.

Human sapoviruses are genetically diverse viruses in the family *Caliciviridae* that can be classified into 19 genogroups, with viruses from genogroups GI, GII, GIV, and GV causing acute gastroenteritis (AGE) in humans. Viruses in these four genogroups can be further divided into 18 genotypes (1–3). We report the near-complete genomic sequences of three sapovirus genotypes (GI.3, GII.2, and GII.6) and two sequences representing a tentative new GII genotype, GII.NA1 (not assigned). These sapoviruses were detected in stool samples from hospitalized patients with AGE in 2006 ($n = 2$), an outpatient child with AGE in 2005 ($n = 1$), and children without diarrhea in 2008 ($n = 2$) in Southeast Kenya (Siaya County). The two GII.NA1 strains were obtained from a child with AGE symptoms in 2005 and from a child without diarrhea enrolled in a prospective case-control study in 2008 (4) and showed 94.9% nucleic acid identity with each other in the VP1 gene and 76% (GII.5) and 73% (GII.3) identity with the closest established sapovirus genotypes.

Viral nucleic acid was extracted from clarified 10% fecal suspensions, and viral metagenomics was performed as described previously (1). Briefly, virus particles were filtered and treated with nucleases, and nucleic acids were extracted using a QIAamp viral RNA minikit (Qiagen, Hilden, Germany). Complementary DNA (cDNA) synthesis by random amplification was performed using a sequence-independent single primer amplification (SISPA) protocol (5). PCR products were purified, and a 300-bp average fragment library was constructed using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA). Samples were sequenced on an Illumina MiSeq system using a MiSeq reagent kit v2 (500 cycles, 2×250 -bp paired-end). A total of 7,444,396 reads were generated, and quality trimming and filtering were performed using a custom bioinformatics pipeline (6, 7). Briefly, raw sequencing reads were filtered to remove host DNA sequences using Bowtie 2 v2.1.0, and the primers and adapters were trimmed using Cutadapt v1.8. After filtering and trimming, 1,956,032 reads (20.6% of the total) remained, 403,533 of which (5.4%) were sapovirus. Complete genomes of human coxsackievirus (Siaya1927, Siaya2158, and Siaya2345), human enterovirus 99 (Siaya0494 and Siaya0506), and “Saffold” virus (Siaya0506) were also found.

Sapovirus genomes were assembled using the *de novo* assembler SPAdes (8) followed by reference mapping and gene annotation using Geneious v11.1.2 (Biomat-

Citation Diez-Valcarce M, Montmayeur A, Tatusov R, Vinjé J. 2019. Near-complete human sapovirus genome sequences from Kenya. *Microbiol Resour Announc* 8:e01602-18. <https://doi.org/10.1128/MRA.01602-18>.

Editor Jelle Matthijnssens, KU Leuven

Copyright © 2019 Diez-Valcarce 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 Jan Vinjé, jvinje@cdc.gov.

Received 2 December 2018

Accepted 21 January 2019

Published 14 February 2019

ters) (9). To compare the assembly results from different samples, we used the *de novo* assembly efficiency metric $UG_{50}\%$ (10). All five samples have a value of $\geq 99.2\%$ or 100.0% , demonstrating direct full-genome assembly.

Strain Hu/KE/2006/GI.3/Siaya1927 (missing 12 nucleotides [nt] of the 5'-untranscribed region [UTR] and 45 nt of open reading frame 1 [ORF1]) is 7,332 nt long. Strain Hu/KE/2006/GII.2/Siaya2158 (missing 13 nt of the 5'-UTR) is 7,382 nt long. Strain Hu/KE/2008/GII.NA1/Siaya0506 (complete genome) is 7,435 nt long. Strain Hu/KE/2005/GII.NA1/Siaya2345 (missing 8 nt of the 5'-UTR and 162 nt in ORF1) is 7,204 nt long. Strain Hu/KE/2008/GII.6/Siaya0494 (missing 13 nt of the 5'-UTR and 18 nt in ORF1) is 7,418 nt long.

Data availability. The genome sequences have been deposited in GenBank under the accession numbers [MG012401](#) and [MH922771](#) to [MH922774](#), and the reads have been deposited in the Sequence Read Archive with the accession numbers [SRR8446701](#) to [SRR8446705](#).

ACKNOWLEDGMENT

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

REFERENCES

- Diez-Valcarce M, Castro CJ, Marine RL, Halasa N, Mayta H, Saito M, Tsaknaridis L, Pan C-Y, Bucardo F, Becker-Dreps S, Lopez MR, Magaña LC, Ng TFF, Vinjé J. 2018. Genetic diversity of human sapovirus across the Americas. *J Clin Virol* 104:65–72. <https://doi.org/10.1016/j.jcv.2018.05.003>.
- Kagning Tsinda E, Malasao R, Furuse Y, Gilman RH, Liu X, Apaza S, Espetia S, Cama V, Oshitani H, Saito M. 2017. Complete coding genome sequences of uncommon GII.8 sapovirus strains identified in diarrhea samples collected from Peruvian children. *Genome Announc* 5:e01137-17. <https://doi.org/10.1128/genomeA.01137-17>.
- Liu X, Jahuir H, Gilman RH, Alva A, Cabrera L, Okamoto M, Xu H, Windle HJ, Kelleher D, Varela M, Verastegui M, Calderon M, Sanchez G, Sarabia V, Ballard SB, Bern C, Mayta H, Crabtree JE, Cama V, Saito M, Oshitani H. 2016. Etiological role and repeated infections of sapovirus among children aged less than two years in a cohort study in a peri-urban community of Peru. *J Clin Microbiol* 54:1598–1604. <https://doi.org/10.1128/JCM.03133-15>.
- Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque ASG, Zaidi AKM, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multi-center Study, GEMS): a prospective, case-control study. *Lancet* 382:209–222. [https://doi.org/10.1016/S0140-6736\(13\)60844-2](https://doi.org/10.1016/S0140-6736(13)60844-2).
- Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E. 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* 86:12161–12175. <https://doi.org/10.1128/JVI.00869-12>.
- Montmayeur AM, Ng TFF, Schmidt A, Zhao K, Magaña 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>.
- 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>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Castro CJ, Ng TFF. 2017. U_{50} : a new metric for measuring assembly output based on non-overlapping, target-specific contigs. *J Comput Biol* 24:1071–1080. <https://doi.org/10.1089/cmb.2017.0013>.