



Near-Complete Genome Sequence of a Human Norovirus GII.1[Pg] Strain Associated with Acute Gastroenteritis, Determined Using Long-Read Sequencing

 Zhihui Yang,^a Mark Mammel,^a Samantha Q. Wales^a

^aDivision of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland, USA

ABSTRACT High-throughput sequencing is one of the approaches used for the detection of foodborne pathogens such as noroviruses. Long-read sequencing has advantages over short-read sequencing in speed, read length, and lower fragmentation bias, which makes it a potential powerful tool for the fast detection and identification of viruses. Using Nanopore sequencing technology, we were able to successfully recover a nearly complete genome sequence of a human norovirus GII.1[Pg] strain in a single long read from a sample from a patient with norovirus gastroenteritis.

Noroviruses are not only the leading cause of foodborne illnesses in the United States but also a primary cause of acute viral gastroenteritis worldwide (1). Noroviruses (genus *Norovirus*, family *Caliciviridae*) can be classified into 10 genogroups (GI to GX) and are further subdivided into at least 48 genotypes. Human noroviruses cluster in genogroups I, II, and IV (2, 3). Due to the high genetic diversity of noroviruses, in addition to PCR-based methods and partial sequencing approaches, high-throughput whole-genome sequencing, which provides nucleotide information on the complete genome, is needed for accurate characterization of these viruses (4, 5). In comparison with short-read sequencing, Nanopore long-read sequencing technologies make it possible to detect and identify viruses faster as well as accurately characterize the genomic complexity with complete coverage of the entire genome by a single read (6). In addition, as a fundamental step during library preparation in short-read sequencing, DNA fragmentation could introduce sequencing bias due to the randomness, incorporation errors, or artifactual indels caused by enzymatic effects (7), and these fragmentation biases can be lowered using Nanopore sequencing technology. Here, we report that using Nanopore sequencing technology and long reads, a nearly complete genome sequence was successfully obtained from a human norovirus GII.1[Pg] strain associated with a case of acute gastroenteritis in the United States in 2016.

A 10% (wt/vol) real-time reverse transcription-PCR (RT-PCR) norovirus-positive stool sample was suspended in phosphate-buffered saline. The supernatant was filtered (via a 4- μ m filter), and viral RNA was extracted using a QIAamp viral RNA minikit (Qiagen) (8). Following RNA concentration (Zymo RNA concentration kit), a library was generated using a cDNA-PCR sequencing kit (Nanopore) and sequenced on the GridION platform (Nanopore). Data analysis was performed with the raw sequencing data using our in-house k-mer tool (9) (github.com/mmammel8/kmer_id) and Long Read Support (beta) module within the CLC Genomics Workbench (CLC GWB version 21). All tools were run with default parameters unless otherwise specified. Out of 117,363 total reads, 105,877 were obtained from norovirus, and the average read length was 1,154 bp (ranging from 151 to 7,572 bp). The longest single read (7,572 bp) was used as a BLAST query against the NCBI nucleotide collection database, and this single read covers 98% of the best reference sequence. With *de novo* assembly using the Long Read Support module, one contig covering the norovirus genome sequence was assembled from the 105,877 reads; the average coverage was 16,242 \times . The

Citation Yang Z, Mammel M, Wales SQ. 2021. Near-complete genome sequence of a human norovirus GII.1[Pg] strain associated with acute gastroenteritis, determined using long-read sequencing. *Microbiol Resour Announc* 10: e00401-21. <https://doi.org/10.1128/MRA.00401-21>.

Editor Jelle Matthijssens, KU Leuven

Copyright © 2021 Yang 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 Zhihui Yang, zhihui.yang@fda.hhs.gov.

Received 20 April 2021

Accepted 11 June 2021

Published 15 July 2021

genome sequence of Hu GII.1[Pg]/Pennsylvania/2016/USA has a length of 7,521 nucleotides (nt) and a GC content of 47.9% and was annotated via the GenBank norovirus submission tool in the NCBI submission portal. It contains (i) three open reading frames (ORFs), ORF1 (5,099 nt), ORF2 (1,608 nt), and ORF3 (774 nt), and (ii) a 3' untranslated region (UTR) 61 nt long. The sequence was genotyped as a norovirus GII.1[Pg] strain using both the Norovirus Genotyping Tool version 2.0 (10) and the CaliciNet lab-based Human Calicivirus Typing Tool (<https://norovirus.ng.philab.cdc.gov>). A full-length genomic sequence BLAST search revealed 96.4% identity with a GII.1[Pg] variant from Indonesia recovered in 2015 (GenBank accession no. [MF668937](https://doi.org/10.1093/jgv.0.001318)).

Data availability. The genome sequence of this GII.1[Pg] strain has been established in GenBank under the accession number [MW854326](https://doi.org/10.1093/jgv.0.001318). The raw sequencing data have been deposited in the SRA under the ViroTrakr project (BioProject accession number [PRJNA396739](https://doi.org/10.1093/jgv.0.001318) and SRA accession number [SRR14119290](https://doi.org/10.1093/jgv.0.001318)).

ACKNOWLEDGMENTS

We thank Mark Craven for his critical reading of the manuscript. The use and application of the deidentified clinical sample in this investigation were conducted under RIHSC (Research Involving Human Subjects Committee-FDA IRB) approval 17-048F.

We declare no conflicts of interest.

REFERENCES

- Lopman B, Zambon M, Brown DW. 2008. The evolution of norovirus, the "gastric flu." *PLoS Med* 5:e42. <https://doi.org/10.1371/journal.pmed.0050042>.
- Chhabra P, de Graaf M, Parra GI, Chan MC-W, Green K, Martella V, Wang Q, White PA, Katayama K, Vennema H, Koopmans MPG, Vinje J. 2019. Updated classification of norovirus genogroups and genotypes. *J Gen Virol* 100:1393–1406. <https://doi.org/10.1099/jgv.0.001318>.
- Chhabra P, de Graaf M, Parra GI, Chan MC-W, Green K, Martella V, Wang Q, White PA, Katayama K, Vennema H, Koopmans MPG, Vinje J. 2020. Corrigendum: updated classification of norovirus genogroups and genotypes. *J Gen Virol* 101:893. <https://doi.org/10.1099/jgv.0.001475>.
- Yang Z, Mammel M, Whitehouse CA, Ngo D, Kulka M. 2018. Inter- and intra-host nucleotide variations in hepatitis A virus in culture and clinical samples detected by next-generation sequencing. *Viruses* 10:619. <https://doi.org/10.3390/v10110619>.
- Silva AJ, Yang Z, Wolfe J, Hirneisen KA, Ruelle SB, Torres A, Williams-Hill D, Kulka M, Hellberg RS. 2021. Application of whole-genome sequencing for norovirus outbreak tracking and surveillance efforts in Orange County, CA. *Food Microbiol* 98:103796. <https://doi.org/10.1016/j.fm.2021.103796>.
- Adikari TN, Riaz N, Siger C, Leung P, Valencia BM, Barton K, Smith MA, Bull RA, Li H, Luciani F, Weeratunga P, Thein T-L, Lim VWX, Leo Y-S, Rajapakse S, Fink K, Lloyd AR, Fernando D, Rodrigo C. 2020. Single molecule, near full-length genome sequencing of dengue virus. *Sci Rep* 10:18196. <https://doi.org/10.1038/s41598-020-75374-1>.
- Tanaka N, Takahara A, Hagio T, Nishiko R, Kanayama J, Gotoh O, Mori S. 2020. Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. *PLoS One* 15:e0227427. <https://doi.org/10.1371/journal.pone.0227427>.
- Yang Z, Mammel MK, Kulka M. 2016. Complete genome sequence of human norovirus GII.4_2006b, a variant of Minerva 2006. *Genome Announc* 4:e01648-15. <https://doi.org/10.1128/genomeA.01648-15>.
- Leonard SR, Mammel MK, Lacher DW, Elkins CA. 2016. Strain-level discrimination of Shiga toxin-producing *Escherichia coli* in spinach using metagenomic sequencing. *PLoS One* 11:e0167870. <https://doi.org/10.1371/journal.pone.0167870>.
- Kroneman A, Vennema H, Deforche K, Avoort HVD, Peñaranda S, Oberste MS, Vinjé J, Koopmans M. 2011. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 51:121–125. <https://doi.org/10.1016/j.jcv.2011.03.006>.