Complete Genome Sequence of a Nontypeable GII Norovirus Detected in Peru

Kentaro Tohma,a Mayuko Saito,b Holger Mayta,c Mirko Zimic,d Cara J. Lepore,a Lauren A. Ford-Siltz,a Robert H. Gilman,e Gabriel I. Parraa

aDivision of Viral Products, Food and Drug Administration, Silver Spring, Maryland, USA 
bDepartment of Virology, Tohoku University Graduate School of Medicine, Sendai, Japan 
cInfectious Diseases Research Laboratory, Department of Cellular and Molecular Sciences, School of Science and Philosophy, Universidad Peruana Cayetano Heredia, Lima, Peru 
dLaboratory of Bioinformatics and Molecular Biology, School of Science and Philosophy, Universidad Peruana Cayetano Heredia, Lima, Peru 
eDepartment of International Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

ABSTRACT Norovirus, a leading cause of acute gastroenteritis in humans, is a highly diverse virus. Here, we report the complete genome sequence of a nontypeable genogroup II (GII) norovirus that was detected in a symptomatic Peruvian child in 2008. This virus showed low nucleotide sequence identities (≈82%) against all known genotypes.

Norovirus is an important cause of acute gastroenteritis. Based on the sequence differences of the major capsid protein (VP1), norovirus is currently divided into seven genogroups (GI to GVII) and more than 30 genotypes (1). Because of the sequence diversity of the RNA-dependent RNA polymerase (RdRp), norovirus typing is based on a dual classification system, which includes polymerase and capsid genotypes (2).

Here, we report the complete genome sequence of a nontypeable GII strain. This strain was detected in a stool sample (PNV06929) collected in 2008 from an 8-month-old Peruvian child with diarrhea (3). The viral RNA was extracted from a 10% stool suspension using the MagMAX viral RNA isolation kit (Ambion, CA), and the full-length norovirus genome was amplified by reverse transcription-PCR, as described by Parra et al. (4). The approximately 7.5-kb PCR product was gel purified using the QIAquick gel extraction kit (Qiagen, CA) and subjected to next-generation sequencing (NGS). The library for NGS was prepared using the Nextera XT DNA library prep kit (Illumina, CA), and the paired-end 2 × 250-bp reads were obtained using the MiSeq system (Illumina). Reads were quality filtered (base quality score, ≥20) and mapped against the reference genomes (Hu/PE/2013/GII.P24-GII.24/Loreto1972, GenBank accession number KY225989; and YURI, GenBank accession number AB083780) to reconstruct a consensus sequence using the HIVE-hexagon (5). A total of 635,163 reads were obtained, covering 7,465 nucleotides (nt) of the viral genome, with an average depth of coverage of 16,664× (range, 230 to 23,398×). The nucleotide sequences of the 5′ untranslated region (UTR) and 3′ UTR of the viral genome were obtained using Sanger sequencing and the 5′/3′ rapid amplification of cDNA ends (RACE) kit (Sigma-Aldrich, MO).

The complete genome of this virus, excluding the poly(A) tail, was 7,525 nt in length and presented three open reading frames (ORFs), ORF1, ORF2, and ORF3, which were 5,091, 1,629, and 765 nt in length, respectively. ORF1 encodes a polyprotein that is cleaved into 6 nonstructural proteins (NS) spanning the sequence as follows: nt 5 to 1000, NS1/2 (p48); nt 1001 to 2098, NS3 (NTPase); nt 2099 to 2620, NS4 (p22); nt 2621...
to 3019, NS5 (VPg); nt 3020 to 3562, NS6 (3C-like protease); and nt 3563 to 5095, NS7 (RdRp). The length of the 5’ UTR was 4 nt, and that of the 3’ UTR was 57 nt, excluding the poly(A) tail. Sequence analyses indicated that the closest strains were GII.P24-GII.24 (Hu/PE/2013/GII.P24-GII.24/Loreto1972; Hu/US/2013/GII.P24-GII.24/EdenPrairieS457, GenBank accession number MG495084; Hu/US/2013/GII.P24-GII.24/EdenPrairieS458, GenBank accession number MG495085) and unassigned strains GII.PNA1-GII.NA1 (Hu/PE/2012/GII.PNA1-GII.NA1/Loreto0959, GenBank accession number MG495077; Hu/PE/2012/GII.PNA1-GII.NA1/Loreto1041, GenBank accession number MG495078) and GII.PNA3-GII.NA3 (Hu/Ni/2005/GII.PNA3-GII.NA3/Leon4509, GenBank accession number KU306738) as of 2 January 2018; all presented 82% nt identity using full-length genome sequences. Phylogenetic analysis of the VP1-encoding sequence was done using MEGA7 (6) and showed no clustering of this virus with any of the GII.24, GII.NA1, or GII.NA3 strains. The RdRp-encoding region of this virus also did not cluster with any known strains. The differences presented with other characterized noroviruses suggest that this virus could form a novel GII genotype. Additional studies conducted in different locations are required to confirm that this could be a new genotype.

Accession number(s). The complete genome sequence of this strain (Hu/GII/PNV06929/2008/PER) has been deposited in GenBank under the accession number MG706448.

ACKNOWLEDGMENTS

We thank the Facility for Biotechnology Resources and the High-Performance Computing Core at U.S. Food and Drug Administration for their support with next-generation sequencing and bioinformatics analysis. We thank the community of San Juan de Miraflores for their collaboration and Asociación Benéfica PRISMA for the fieldwork. We also thank Jan Vinjé from the Centers for Disease Control and Prevention for his help with genotyping analyses.

Financial support for this work was provided by U.S. Food and Drug Administration intramural funds. K.T. and C.J.L. are recipients of a CBER/FDA-sponsored Oak Ridge Institute for Science and Education (ORISE) fellowship.

The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Vinjé J. 2015. Advances in laboratory methods for detection and typing of norovirus. J Clin Microbiol 53:373–381. https://doi.org/10.1128/JCM.01535-14.

2. 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.

3. Saito M, Goel-Apaza S, Espetia S, Velasquez D, Cabrera L, Loli S, Crabtree JE, Black RE, Kosek M, Checkley W, Zimic M, Bern C, Cama V, Gilman RH, Xiao L, Kelleher D, Windle HJ, van Doorn LJ, Varela M, Verastegui M, Calderon M, Alva A, Roman K. Norovirus Working Group in Peru. 2014. Multiple norovirus infections in a birth cohort in a Peruvian periurban community. Clin Infect Dis 58:483–491. https://doi.org/10.1093/cid/cit683.

4. Parra GI, Squires RB, Karangwa CK, Johnson JA, Lepore CJ, Sosnovtsev SV, Green KY. 2017. Static and evolving norovirus genotypes: implications for epidemiology and immunity. PLoS Pathog 13:e1006136. https://doi.org/10.1371/journal.ppat.1006136.

5. Santana-Quintero L, Dingerdissen H, Thierry-Mieg J, Mazumder R, Simonyan V. 2014. HIVE-hexagon: high-performance, parallelized sequence alignment for next-generation sequencing data analysis. PLoS One 9:e99033. https://doi.org/10.1371/journal.pone.0099033.

6. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.