Genome Sequence of an Unknown Subtype of Hepatitis C Virus Genotype 6

Another Piece for the Taxonomic Puzzle

Pedersen, Martin Schou; Mollerup, Sarah; Nielsen, Lone Gilmor; Jenssen, Håvard; Bukh, Jens; Schønning, Kristian

Published in:
Microbiology Resource Announcements

DOI:
10.1128/MRA.01030-19

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Pedersen, M. S., Mollerup, S., Nielsen, L. G., Jenssen, H., Bukh, J., & Schønning, K. (2019). Genome Sequence of an Unknown Subtype of Hepatitis C Virus Genotype 6: Another Piece for the Taxonomic Puzzle. *Microbiology Resource Announcements*, *8*, [e01030-19]. https://doi.org/10.1128/MRA.01030-19

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequence of an Unknown Subtype of Hepatitis C Virus Genotype 6: Another Piece for the Taxonomic Puzzle

Martin Schou Pedersen,[a,b,c,d] Sarah Mollerup,[a] Lone Gilmor Nielsen,[a] Håvard Jenssen,[b] Jens Bukh,[c,d] Kristian Schønning[a,e]

[a]Department of Clinical Microbiology, Copenhagen University Hospital, Hvidovre, Denmark
[b]Department of Science and Environment, Roskilde University, Roskilde, Denmark
[c]Copenhagen Hepatitis C Program (CO-HEP), Department of Infectious Diseases, Copenhagen University Hospital, Hvidovre, Denmark
[d]Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[e]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

ABSTRACT The surveillance and correct subtyping of hepatitis C virus strains require available and up-to-date publicly available reference genomes. Here, we present the complete open reading frame sequence of a hepatitis C virus genotype 6 strain of an unknown subtype that was discovered during routine subtyping of patients in the clinic.

Hepatitis C virus (HCV) is a worldwide pathogen that belongs to the genus *Hepacivirus* within the *Flaviviridae* family. The viral genome is positive-sense, single-stranded RNA, approximately 9,600 nucleotides long, with a single open reading frame (ORF) about 9,000 nucleotides long (1). There are 8 recognized main variants of HCV (genotypes 1 to 8) with up to 35% nucleotide divergence and 90 accepted subtypes deviating up to 25% (2). Genotype 6, the most diverse of the genotypes, is most commonly observed in Asia (3) and continues to be expanded with novel subtype sequences (4). The International Committee on Taxonomy of Viruses (ICTV) requires 3 independent isolates to accept a new subtype (2). Here, we present the HCV ORF sequence of a yet-to-be-defined subtype identified from a patient sample (HVH-HCV334) in January 2019 in Copenhagen, Denmark, during routine analyses.

The sample had a viral load of 7.51 log IU/ml, as measured by the Aptima HCV Quant Dx assay (5). RNA was extracted with the ZR viral RNA kit (Zymo Research) as described (6) and depleted for human rRNA with the NEBNext rRNA depletion kit (New England BioLabs). RNA sequencing (RNA-seq) libraries were prepared with the NEBNext Ultra II directional RNA library prep kit (New England BioLabs) in half the standard reaction volume suggested by the manufacturer. Sequencing was performed with 2 × 150-bp reads on a MiSeq instrument (Illumina). All software was used with default parameters unless specified. Reads (~4.5 million) were trimmed and quality filtered with fastp v.0.12.2 (7) to retain a Phred quality of >20 and reads of >50 bp. Filtered reads (~4 million) were mapped to the human genome hg38 (GenBank accession no. GCA_000001405.27) with Bowtie2 v.2.3.4.1 (8), and unmapped reads (~2.5 million) were sorted by SAMtools v.1.9 (9), extracted with BEDtools v.2.26.0 (10), and assembled with VICUNA v.1.3 (11). The HCV ORF was identified and annotated with Geneious v.10.2.3 (12) based on reference strain H77 (GenBank accession no. NC_004102). The ORF was 9,069 nucleotides long, without premature stop codons, and annotation identified 3 structural proteins and 7 nonstructural proteins flanked by incomplete 5′ and 3′ untranslated regions. No recombination or subgenomic deletion variants were observed by previously described methods (6). The Geneious statistics function reported a depth of coverage of ~24,500 and a G+C content of 56%. All

Citation Pedersen MS, Mollerup S, Nielsen LG, Jenssen H, Bukh J, Schønning K. 2019. Genome sequence of an unknown subtype of hepatitis C virus genotype 6: another piece for the taxonomic puzzle. Microbiol Resour Announc 8:e01030-19. https://doi.org/10.1128/MRA.01030-19.

Editor Kenneth M. Stedman, Portland State University

Address correspondence to Kristian Schønning, Kristian.schoenning@regionh.dk.

Received 30 August 2019
Accepted 20 September 2019
Published 17 October 2019

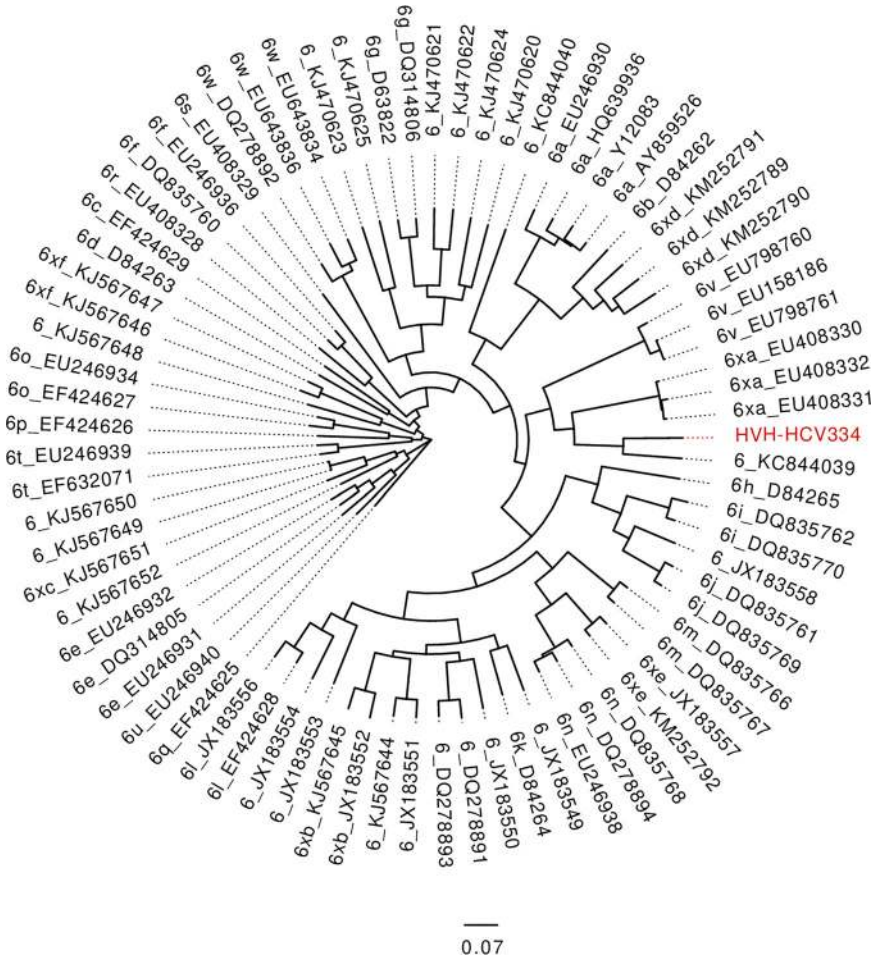

FIG 1 Phylogenetic tree with International Committee on Taxonomy of Viruses (ICTV) genotype 6 reference samples and genotype 6 samples without a designated subtype according to the ICTV classification. Sequences identified as genotype 6 according to the ICTV classification were obtained from NCBI GenBank. Samples were aligned with MUSCLE v.3.8.1551; a maximum likelihood phylogenetic tree was created with FastTree v.2.1.5 and visualized in FigTree v.1.4.3. Branch labels show the designated subtypes and NCBI GenBank accession numbers for the individual samples. HVH-HCV334 is colored red.

official genotype 6 references and sequences without subtype assignment according to the International Committee on Taxonomy of Viruses (ICTV) classification from May 2019 (2) and HVH-HCV334 were aligned with MUSCLE v.3.8.1551 (13); a maximum likelihood phylogenetic tree was created with FastTree v.2.1.5 (14) and visualized in FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). As seen in Fig. 1, HVH-HCV334 was located close to the untyped sample with GenBank accession no. KC844039 (15) from China and had 81% pairwise nucleotide identity across the ORF. Possible resistance toward ombitasvir, and thus potentially other NS5A inhibitors, was predicted, by HCV GLUE v.0.1.58 (16), due to the resistance-associated amino acids M28 (99%) and S93 (99%) in the NS5A protein (17, 18).

This new genotype 6 genome sequence is important for accurate characterization of HCV for surveillance and prior to antiviral treatment with nonpangenotypic regimens.

Data availability. The sequencing reads have been deposited at NCBI under BioProject no. PRJNA557264, and the genome sequence has been deposited in GenBank under accession no. MN240359.

ACKNOWLEDGMENTS

This work was supported by a Ph.D. stipend from Roskilde University (M.S.P.) and by grants from The Capital Region of Denmark's Research Foundation (J.B.), the Novo

Nordisk Foundation (J.B.), and the Danish Research Council (J.B.), including an advanced Sapere Aude grant (J.B.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Bukh J. 2016. The history of hepatitis C virus (HCV): basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. J Hepatol 65:S2–S21. https://doi.org/10.1016/j.jhep.2016.07.035.
2. Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, Muerhoff S, Pletnev A, Rico-Hesse R, Smith DB, Stapleton JT, ICTV Report Consortium. 2017. ICTV virus taxonomy profile: Flaviviridae. J Gen Virol 98:2–3. https://doi.org/10.1099/jgv.0.000672.
3. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, Barnes E. 2015. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology 61:77–87. https://doi.org/10.1002/hep.27259.
4. Xu R, Yu Y, Leitch ECM, Wang M, Huang K, Huang J, Tang X, Liao Q, Song D, Shan Z, Li C, Mclauchlan J, Rong X. 2019. HCV genotype 6 prevalence, spontaneous clearance and diversity among elderly members of the Li ethnic minority in Baisha County, China. J Viral Hepat 26:529–540. https://doi.org/10.1111/jvh.13062.
5. Schønning K, Pedersen MS, Johansen K, Landt B, Nielsen LG, Weis N, Westh H. 2017. Analytical and clinical performance of the Hologic Aptima HCV Quant Dx assay for the quantification of HCV RNA in plasma samples. J Virol Methods 248:159–165. https://doi.org/10.1016/j.jviromet.2017.07.006.
6. Pedersen MS, Fahnøe U, Hansen TA, Pedersen AG, Jenssen H, Bukh J, Schønning K. 2018. A near full-length open reading frame next generation sequencing assay for genotyping and identification of resistance-associated variants in hepatitis C virus. J Clin Virol 105:49–56. https://doi.org/10.1016/j.jcv.2018.05.012.
7. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
10. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841. https://doi.org/10.1093/bioinformatics/btq033.
11. Yang X, Charlebois P, Gnerre S, Coole MG, Lennon NJ, Levin JZ, Qu J, Ryan EM, Zody MC, Henn MR. 2012. De novo assembly of highly diverse viral populations. BMC Genomics 13:475. https://doi.org/10.1186/1471-2164-13-475.
12. 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.
13. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
14. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641–1650. https://doi.org/10.1093/molbev/msp077.
15. Xu R, Tong W, Gu L, Li C, Fu Y, Lu L. 2013. A panel of 16 full-length HCV genomes was characterized in China belonging to genotypes 1–6 including subtype 2f and two novel genotype 6 variants. Infect Genet Evol 20:225–229. https://doi.org/10.1016/j.meegid.2013.08.014.
16. Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ. 2018. GLUE: a flexible software system for virus sequence data. BMC Bioinformatics 19:532. https://doi.org/10.1186/s12859-018-2459-9.
17. Gottwein JM, Pham LV, Mikkelsen LS, Ghanem L, Ramirez S, Scheel TKH, Carlsen THR, Bukh J. 2018. Efficacy of NS5A inhibitors against hepatitis C virus genotypes 1–7 and escape variants. Gastroenterology 154:1435–1448. https://doi.org/10.1053/j.gastro.2017.12.015.
18. Pham LV, Ramirez S, Gottwein JM, Fahnøe U, Li Y-P, Pedersen J, Bukh J. 2018. HCV genotype 6a escape from and resistance to velpatasvir, pibrentasvir, and sofosbuvir in robust infectious cell culture models. Gastroenterology 154:2194. https://doi.org/10.1053/j.gastro.2018.02.017.