

Coding-complete genome of human alphaherpesvirus 1 isolated from a case of fulminant hepatitis

Natalie Viljoen,^{1,2,3} Felicity Burt,^{2,4} Jacqueline Weyer^{1,3,5}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the coding-complete genome sequence of human alphaherpesvirus 1 (HHV1) isolated from a previously healthy 64-year-old male with fulminant hepatitis, a rare presentation of a common viral agent. The sequence is highly similar to previously described HHV1 sequences. Additional sequence data for fulminant hepatitis cases are required.

KEYWORDS hepatitis, human alphaherpes virus 1, herpesviruses

Human alphaherpesvirus 1 (HHV1) belongs to the family *Herpesviridae* and is typically associated with mucocutaneous vesicular lesions, and disseminated disease is reported to be rare (1). HHV1 DNA was detected in the serum of a 64-year-old, previously healthy, male by qPCR confirming a diagnosis of fulminant herpetic hepatitis. Ethics approval was obtained from the Human Research Ethics Committee (Medical) at the University of the Witwatersrand (reference number: R14/49). Virus was isolated from serum (stored at -80°C) in Vero cells (ATCC CRL-1586), which were monitored for characteristic cytopathic effect (2). Cells were harvested after the second passage and concentrated from cleared supernatant using an Ultra-0.5 centrifugal filter with a 100-kDa cutoff (Amicon) in preparation for downstream processing for short- and long-read sequencing. For short-read sequencing, DNA was extracted using the Quick-DNA/RNA Viral Kit (Zymo Research), quantified using a Qubit fluorometer (Thermo Fisher Scientific), paired-end libraries prepared using the Nextera XT Library Prep kit (Illumina) for sequencing on the MiSeq (Illumina) platform (2×300 bp). For long-read sequencing, virus was purified through a 10% sucrose cushion, extracted using a genomic tip 20 /G (Qiagen), quantified using a Qubit fluorometer (Thermo Fisher Scientific), libraries prepared using the SMRTbell Express Template Prep kit 2.0 (Pacific BioSciences), DNA sheared using the Megaruptor 3 (Diagenode), and size selected using Ampure PB beads (Pacific BioSciences) for sequencing on the Sequel II system (Pacific BioSciences) using the Sequel II binding kit 2.2 and sequencing plate V2 (Pacific BioSciences).

FASTQ files were uploaded to the Galaxy Web platform and data analyzed using the server at <http://usegalaxy.org> (3). All tools were run using the default parameters, and short-read data were run for paired-end reads unless otherwise noted. Reads were quality trimmed (qualified quality phred, 20; length required for short reads, 40; length required for long reads, 100) using fastp V0.19.5 (4). For *de novo* assembly, reads were filtered against the reference sequence, assembled using MEGAHIT V1.2.9 (5) and evaluated using QUAST V5.2.0 (6). In addition, short and long reads were mapped on the reference sequence (GenBank accession number: NC001806.1 available at https://www.ncbi.nlm.nih.gov/nuccore/NC_001806) using Bowtie2 V2.4.5 (7) and Minimap2 V2.24 (8), respectively, and consensus sequence determined using iVar V1.3.1 (Minimum depth, 2) (9). Draft genomes obtained were aligned using MAFFT (10) and were in agreement. The most complete draft genome was obtained by long-read

Editor Kenneth M. Stedman, Portland State University, Portland, Oregon, USA

Address correspondence to Jacqueline Weyer, jacquelinew@nicd.ac.za.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 11 May 2023

Accepted 3 August 2023

Published 25 September 2023

Copyright © 2023 Viljoen 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/).

TABLE 1 Genome assembly metrics for short- and long-read data

	Short-read data (MiSeq, Illumina)	Long-read data (Sequel II, PacBio)
Total reads	2.8 million	14.1 thousand
Average read length	224 bp	3,675 bp
Genome coverage distribution for reference-based assembly	151,379 bp	152,159 bp
Average sequencing depth across genome	485×	12×
Genome coverage distribution with <i>de novo</i> assembly	134,520 bp	136,968 bp
Number of contigs for <i>de novo</i> assembly	11	9
N ₅₀ for <i>de novo</i> assembly	44,644 bp	45,163 bp

reference-based mapping, which was manually curated at ambiguous positions using Integrative Genomics Viewer V2.13.0 (11). The sequencing depth was determined using the SAMtools suite (12).

The genome is 152,063 bp in length (GC content 68.22%). Genome assembly metrics are provided in Table 1. The genome contains a 110 bp gap (position 117,349–117,458) in an inverted repeat region, which could not be resolved. Genome annotation was performed using NCBI ORF finder (13), DIAMOND (14), and the NCBI protein database (15). BLAST (16) analysis suggested that this sequence had a nucleotide identity of 98.75% with the reference sequence. Additional sequence data from HHV1 fulminant hepatitis cases are required to aid in studying potential virus-related factors that may play a role in the development of severe disease.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation (NRF) of South Africa [116729 to N.V.], the NRF South African Research Chairs Initiative [U98346 to F.J.B], and operational funding at the NICD/NHLS. The authors acknowledge the technical support offered by the Core Sequencing Facility of the NICD/NHLS in performing MiSeq sequencing.

The funders had no role in the study design, data collection and analysis, decision to publish or the preparation of the manuscript.

AUTHOR AFFILIATIONS

¹Center for Emerging Zoonotic and Parasitic Diseases, National Institute for Communicable Disease of the National Health Laboratory Service, Sandringham, South Africa

²Division of Virology, University of the Free State, Bloemfontein, South Africa

³Centre for Viral Zoonoses, Department of Medical Virology, University of Pretoria, Pretoria, South Africa

⁴Division of Virology, National Health Laboratory Service, Universitas, Bloemfontein, South Africa

⁵Department of Microbiology and Infectious Diseases, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa

AUTHOR ORCIDs

Natalie Viljoen  <http://orcid.org/0000-0002-9959-2004>

Jacqueline Weyer  <http://orcid.org/0000-0001-9471-2890>

FUNDING

Funder	Grant(s)	Author(s)
National Research Foundation (NRF)	116729	Natalie Viljoen

Funder	Grant(s)	Author(s)
National Research Foundation (NRF)	U98346	Felicity Jane Burt
National Institute for Communicable Diseases	NA	Jacqueline Weyer

AUTHOR CONTRIBUTIONS

Natalie Viljoen, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | Felicity Burt, Conceptualization, Resources.

DATA AVAILABILITY

This sequence has been deposited in GenBank under the accession number [OQ102003.1](https://www.ncbi.nlm.nih.gov/nuccore/OQ102003.1) (<https://www.ncbi.nlm.nih.gov/nuccore/OQ102003.1>). The version described here is the first version. The raw reads were deposited in the NCBI Sequence Read Archive (SRA) under the accession number [PRJNA913943](https://www.ncbi.nlm.nih.gov/sra/PRJNA913943) (<https://www.ncbi.nlm.nih.gov/sra/PRJNA913943>) and [SRX18774967](https://www.ncbi.nlm.nih.gov/sra/SRX18774967) and <https://www.ncbi.nlm.nih.gov/sra/SRX18774968>.

REFERENCES

- Norvell JP, Blei AT, Jovanovic BD, Levitsky J. 2007. Herpes simplex virus hepatitis: an analysis of the published literature and institutional cases. *Liver Transpl* 13:1428–1434. <https://doi.org/10.1002/lt.21250>
- World Health Organization. 2017. Use of cell culture in Virology for developing countries in the South-East Asia region
- Schatz MC, Philippakis AA, Afgan E, Banks E, Carey VJ, Carroll RJ, Culotti A, Elliott K, Goecks J, Grossman RL, Hall IM, Hansen KD, Lawson J, Leek JT, Luria AO, Mosher S, Morgan M, Nekrutenko A, O'Connor BD, Osborn K, Paten B, Patterson C, Tan FJ, Taylor CO, Vessio J, Waldron L, Wang T, Wuichet K. 2022. The Galaxy platform for accessible, reproducible and collaborative BIOMEDICAL analyses. *Nucleic Acids Res* 50 2:W345–W351. <https://doi.org/10.1016/j.xgen.2021.100085>
- 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>
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
- Langmead B, Salzberg SL. 2012. fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, Manfredini R, Luppi M, Trenti T, Tagliafico E. 2021. iVar, an interpretation-oriented tool to manage the update and revision of variant annotation and classification. *Genes* 12:384. <https://doi.org/10.3390/genes12030384>
- 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>
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L. 2003. Database resources of the national center for biotechnology. *Nucleic Acids Res* 31:28–33. <https://doi.org/10.1093/nar/gkg033>
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50:D20–D26. <https://doi.org/10.1093/nar/gkab1112>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. J Mol Biolbasic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)