



Complete Genome Sequence of Herpes Simplex Virus 1 Strain McKrae

Xiaoli Jiao,^a  Hongyan Sui,^a Christopher Lyons,^b Bao Tran,^b Brad T. Sherman,^a Tomozumi Imamichi^a

^aLaboratory of Human Retrovirology and Immunoinformatics, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

^bSequencing Facility, Center for Cancer Research, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

ABSTRACT Herpes simplex virus 1 (HSV-1) strain McKrae is highly virulent and relatively neuroinvasive in animal models compared with other wild-type HSV-1 strains. To identify the genetic determinants that lead to the unique phenotypes of the McKrae strain, we sequenced its genome with PacBio single-molecule real-time (SMRT) technology and resolved the complete sequence.

Herpes simplex virus 1 (HSV-1) is a member of the *Alphaherpesvirinae* subfamily of the *Herpesviridae* family (1). HSV-1 strain McKrae has a double-stranded DNA genome (~152 kb) originally isolated from a patient with herpes simplex keratitis (2). It undergoes spontaneous or induced reactivation at a higher frequency than do other HSV-1 strains (e.g., strain KOS) and is among the most virulent HSV-1 strains in mice and rabbits (3–7). Ocular infection by HSV-1 McKrae causes fatal encephalitis in approximately 50% of animals (8). The two existing genome sequences of McKrae strains (GenBank accession numbers [JX142173](#) and [JQ730035](#)) were sequenced with Illumina and Roche 454 platforms, respectively, and ~3.5-kb and ~1.5-kb genomic regions were undetermined in the assemblies (8, 9).

We obtained HSV-1 strain McKrae from Jeffrey I. Cohen (National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD). The virus was propagated in Vero cells (ATCC, Gaithersburg, MD) (10). The genomic DNA was isolated using the PureLink viral DNA minikit (Invitrogen, Carlsbad, CA), following the manufacturer's protocol. The viral DNA shearing was carried out using an M220 sonicator (Covaris, Woburn, MA). Size selection and subsequent cleanup were performed using a combination of BluePippin (Sage Science, Beverly, MA) with a miniTUBE Red (Covaris) and AMPure PB beads (PacBio, Menlo Park, CA), respectively. After shearing, bands of ~5 kb in length were used for library construction using the SMRTbell template prep kit 1.0 (PacBio) with two alterations, namely, an A-tailing step and overhang adapter ligation, as opposed to blunt adapter ligation (11). The constructed libraries were then sequenced using PacBio long-read sequencing on a Sequel system using v3.0 chemistry, with a 5 pM loading concentration and a 10-hour runtime.

The HSV-1 genome (Fig. 1a) is GC rich and contains many repeat elements, including two copies of a long inverted repeat (RL) (~8.75 kb each), two copies of a short inverted repeat (RS) (~6.25 kb each) (8), and three copies of an "a" sequence, which is repeated at both ends of the genome and at the internal long-short (L-S) junction (12). Figure 1b shows a loop formed by these repeat elements which causes a complication for *de novo* assembly to resolve a full-length linear genome sequence, even with long reads. We obtained five contigs (max length, 126,990 bp; N_{50} length, 33,615 bp) from the PacBio Hierarchical Genome Assembly Process v4 (HGAP4) (13), with default settings, except with "seed length cutoff" set to 15 kb and "minimum required alignment length" set to 500 bp with an input genome length of 160 kb. The alignment of the five contigs (Fig. 1a) against the genome sequence of HSV-1 reference strain 17 (GenBank accession

Citation Jiao X, Sui H, Lyons C, Tran B, Sherman BT, Imamichi T. 2019. Complete genome sequence of herpes simplex virus 1 strain McKrae. *Microbiol Resour Announc* 8:e00993-19. <https://doi.org/10.1128/MRA.00993-19>.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2019 Jiao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Tomozumi Imamichi, timamichi@mail.nih.gov.

X.J. and H.S. contributed equally to this work.

Received 14 August 2019

Accepted 30 August 2019

Published 26 September 2019

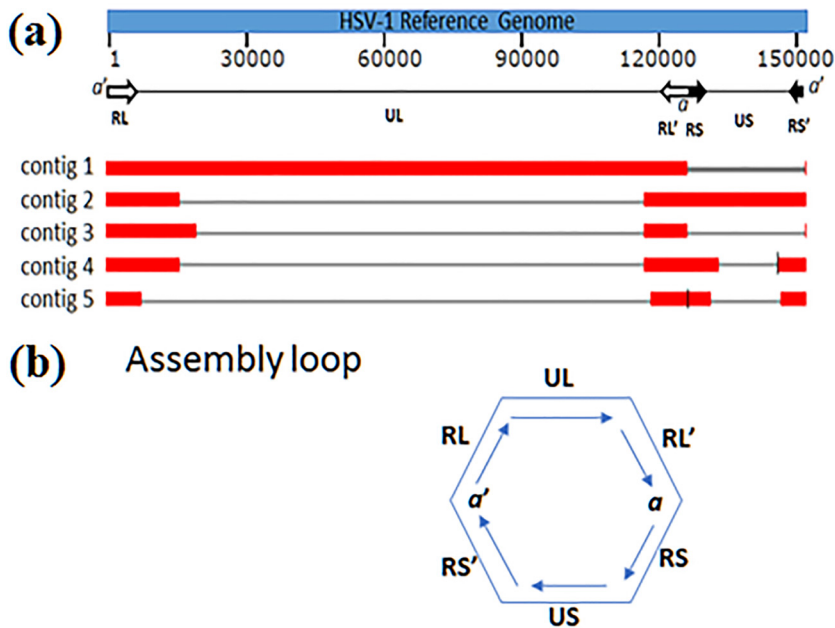


FIG 1 HSV-1 genome and alignments of five contigs. (a) The alignments of the five contigs against the reference genome HSV-1 strain 17. The nucleotide lengths of contigs 1, 2, 3, 4, and 5 are 126,990 bp, 44,947 bp, 20,043 bp, 33,615 bp, and 13,330 bp, respectively. The HSV-1 genome is composed of a unique long (UL) region (106.5 kb), a unique short (US) region (13.5 kb), two copies of a long inverted repeat (RL and RL'), and two copies of a short inverted repeat (RS and RS') (6.25 kb each). The "a" sequence is repeated at both ends of the genome and at the internal L-S junction. There are multiple alignments for one contig due to the presence of the repeated elements in the reference genome. The gaps showing between the alignments for a contig are not deletions. (b) Assembly loop formed by the genome elements.

number [JN555585](#)) by BLASR (v2.0.0) (14), with default settings, shows that contig 1 (126,990 bp) and contig 2 (44,947 bp) jointly covered the full length of the reference genome from the 5' to 3' ends, and we merged them into a draft sequence. The genome termini were manually determined by the *a* sequence (303 bp) bracketed by two DR1 elements, which also had an inverted repeat in the L-S junction (15). The draft sequence was then refined by the PacBio genome resequencing pipeline (SMRT Analysis v6.0), with default settings (16). A total of 4,834,368 subreads (mean length, 2,183; maximum length, 26,927; total bases, 10,553,016,021) were mapped to the final resolved genome sequence. The average depth per base pair is 64,245 reads, a coverage that ensures high-quality consensus base calling (>99.99% accuracy) (13).

The finished length of the complete genome is 152,361 bases, with a G+C content of 68.21%. It resolves ~3.7 kb of a total of 11 gaps in the two GenBank records (accession numbers [JX142173](#) and [JQ730035](#)), including an ~100-bp gap in the PQ repeat region of UL36. UL56 contains a single base pair insertion frameshift (amino acid 97) that results in a divergent and truncated C terminus, which was reported in [JQ730035](#) (8) but not found in [JX142173](#) (9). Only one coding mutation (L61P) in UL53 was found in the assembly due to a single nucleotide variation at 112,361 (T < C), but it was not detected in [JX142173](#) and [JQ730035](#). The high concordance (>99.8% identity) between the assembly and the two GenBank records in the nongapped regions indicates the high quality of the assembly.

Data availability. The complete genome sequence has been deposited at GenBank under the accession number [MN136524](#), with SRA accession number [SRR9719184](#). The annotations for the genes and coding sequences were transferred from HSV-1 strain 17 to the McKrae strain based on sequence homology (17, 18).

ACKNOWLEDGMENTS

We thank H. Clifford Lane for supporting this project and Jeffrey I. Cohen for providing HSV-1 strain McKrae.

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. This research was supported, in part, by the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

- Whitley RJ, Kimberlin DW, Roizman B. 1998. Herpes simplex viruses. *Clin Infect Dis* 26:541–555. <https://doi.org/10.1086/514600>.
- Williams LE, Nesburn AB, Kaufman HE. 1965. Experimental induction of disciform keratitis. *Arch Ophthalmol* 73:112–114. <https://doi.org/10.1001/archophth.1965.00970030114023>.
- Wang H, Davido DJ, Morrison LA. 2013. HSV-1 strain McKrae is more neuroinvasive than HSV-1 KOS after corneal or vaginal inoculation in mice. *Virus Res* 173:436–440. <https://doi.org/10.1016/j.virusres.2013.01.001>.
- Perng GC, Mott KR, Osorio N, Yukht A, Salina S, Nguyen QH, Nesburn AB, Wechsler SL. 2002. Herpes simplex virus type 1 mutants containing the KOS strain ICP34.5 gene in place of the McKrae ICP34.5 gene have McKrae-like spontaneous reactivation but non-McKrae-like virulence. *J Gen Virol* 83:2933–2942. <https://doi.org/10.1099/0022-1317-83-12-2933>.
- Sawtell NM, Poon DK, Tansky CS, Thompson RL. 1998. The latent herpes simplex virus type 1 genome copy number in individual neurons is virus strain specific and correlates with reactivation. *J Virol* 72:5343–5350.
- Strelow LI, Laycock KA, Jun PY, Rader KA, Brady RH, Miller JK, Pepose JS, Leib DA. 1994. A structural and functional comparison of the latency-associated transcript promoters of herpes-simplex virus type 1 strains KOS and McKrae. *J Gen Virol* 75:2475–2480. <https://doi.org/10.1099/0022-1317-75-9-2475>.
- Hill JM, Rayfield MA, Haruta Y. 1987. Strain specificity of spontaneous and adrenergically induced HSV-1 ocular reactivation in latently infected rabbits. *Curr Eye Res* 6:91–97. <https://doi.org/10.3109/02713688709020074>.
- Watson G, Xu W, Reed A, Babra B, Putman T, Wick E, Wechsler SL, Rohrmann GF, Jin L. 2012. Sequence and comparative analysis of the genome of HSV-1 strain McKrae. *Virology* 433:528–537. <https://doi.org/10.1016/j.virol.2012.08.043>.
- Macdonald SJ, Mostafa HH, Morrison LA, Davido DJ. 2012. Genome sequence of herpes simplex virus 1 strain McKrae. *J Virol* 86:9540–9541. <https://doi.org/10.1128/JVI.01469-12>.
- Harper DR. 2000. Determination of viral infectivity, p 119–127. *In* Kinchington D, Schinazi RF (ed), *Antiviral methods and protocols*. Humana Press, Totowa, NJ.
- Jiao X, Sui H, Lyons C, Tran B, Sherman BT, Imamichi T. 2019. Complete genome sequence of herpes simplex virus 1 strain MacIntyre. *Microbiol Resour Announc* 8:e00895-19. <https://doi.org/10.1128/MRA.00895-19>.
- Umene K. 1993. Herpes-simplex virus type 1 variant a sequence generated by recombination and breakage of the a sequence in defined regions, including the one involved in recombination. *J Virol* 67:5685–5691.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. <https://doi.org/10.1186/1471-2105-13-238>.
- Umene K, Ohashi S, Yoshida M, Fukumaki Y. 2008. Diversity of the a sequence of herpes simplex virus type 1 developed during evolution. *J Gen Virol* 89:841–852. <https://doi.org/10.1099/vir.0.83467-0>.
- Morse AM, Calabro KR, Fear JM, Bloom DC, McIntyre LM. 2017. Reliable detection of herpes simplex virus sequence variation by high-throughput resequencing. *Viruses* 9:226. <https://doi.org/10.3390/v9080226>.
- McGeoch DJ, Dalrymple MA, Davison AJ, Dolan A, Frame MC, McNab D, Perry LJ, Scott JE, Taylor P. 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol* 69:1531–1574. <https://doi.org/10.1099/0022-1317-69-7-1531>.
- McGeoch DJ, Dolan A, Donald S, Brauer D. 1986. Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res* 14:1727–1745. <https://doi.org/10.1093/nar/14.4.1727>.