JVI Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of Herpes Simplex Virus 1 Strain McKrae

Stuart J. Macdonald,[a] Heba H. Mostafa,[a] Lynda A. Morrison,[b] and David J. Davido[a]

Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, USA,[a] and Department of Molecular Microbiology and Immunology, Saint Louis University School of Medicine, St. Louis, Missouri, USA[b]

The herpes simplex virus 1 (HSV-1) strain McKrae is highly virulent compared to other wild-type strains of HSV-1. To help us better understand the genetic determinants that lead to differences in the pathogenicity of McKrae and other HSV-1 strains, we sequenced its genome. Comparing the sequence of McKrae's genome to that of strain 17 revealed that the genomes differ by at least 752 single nucleotide polymorphisms (SNPs) and 86 insertion/deletion events (indels). Although the majority of these polymorphisms reside in noncoding regions, 241 SNPs and 10 indels alter the protein-coding sequences of 58 open reading frames. Some of these variations are expected to contribute to the pathogenic phenotype of McKrae.

Herpes simplex virus 1 (HSV-1) is a double-stranded DNA, enveloped virus of the *Herpesviridae* family. HSV-1 infections can be asymptomatic or cause cold sores, corneal ulcerations, or life-threatening encephalitis (9). Studies examining a variety of HSV-1 strains show a broad spectrum of virulent and pathogenic phenotypes. Notably, the HSV-1 strain McKrae reactivates at a high frequency and is more neurovirulent in animal models than other strains of HSV-1 (4, 7, 8, 10, 11). Because viral determinants that contribute to McKrae's *in vivo* phenotypes are largely unknown, we sequenced its genome.

McKrae genomic DNA was isolated from infected Vero (African green monkey kidney) cells (2) and used to construct an unpaired Illumina library. All raw 42-bp reads were assembled against rhesus macaque and human genomes using Bowtie (5) to eliminate any reads derived from the host DNA. Low-quality sequence reads were removed using the SolexaQA Perl scripts (3). The remaining 6,528,420 high-quality reads were assembled with the Velvet *de novo* assembler (13). Resulting contigs of >100 bp were semiautomatically assembled against the reference HSV-1 strain 17 genome (GenBank accession number NC_001806) using SeqMan Pro (DNAStar, Inc.). Since HSV-1 includes two sets of inverted repeat regions, TRL/IRL and IRS/TRS, contigs assembling into one of the repeat units were reverse complemented and placed into the other repeat unit. We validated this new genome, particularly the sequence between *de novo* contigs, by using Bowtie to align the set of filtered reads against the assembly. A comparison of our McKrae genome to six Sanger-sequenced McKrae genes, UL1, UL20, UL22, UL27, UL44, and UL53 (2a), revealed six differences out of 10,314 comparable bases, five of which were in the UL27 gene. Thus, we have high confidence in the sequence of our genome assembly.

We sequenced the McKrae genome to an average depth of 1,700 reads per base pair, a depth which is comparable to that of other short-read-based HSV-1 genomes (12). The final genome is 152,201 bp and has 10 gaps totaling 3,570 bp. All gaps, including two in the unique long (UL) region and several at variable-number tandem-repeat (VNTR) regions, are clearly marked in the GenBank annotation, and the missing sequences are given as strings of nucleotides that correspond to the lengths of the regions in strain 17. We used the rapid annotation transfer tool (RATT) (6) to transfer coding region annotations from strain 17 to the McKrae genome.

Genome alignment of strain 17 and McKrae (1), followed by application of a series of custom Perl and R scripts, identified 752 single nucleotide polymorphisms (SNPs) and 86 insertion/deletion events (indels) distinguishing these two strains. Around one-third of the SNPs and virtually all of the indels reside in noncoding regions. However, 241 SNPs and 10 indels change the protein-coding sequences of 58 of the 77 open reading frames between strains 17 and McKrae, likely contributing to differences in their virulence.

Nucleotide sequence accession number. The HSV-1 strain McKrae genome in GenBank has the accession number JX142173.

ACKNOWLEDGMENTS

We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with sequencing.

The Genome Technology Access Center is partially supported by NCI Cancer Center Support grant number P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA grant number UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This research was supported by NIH grants R01AI72357 (D.J.D.), R21EY019739 (L.A.M. and D.J.D.), R01RR024862 and R01GM085260 (S.J.M.), and R21NS070417 (S.J.M. and E. Lundquist).

Support from the Department of Molecular Biosciences (a Linux workstation), obtained with the assistance of R. Cohen, is greatly appreciated.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. **Bradley RK, et al.** 2009. Fast statistical alignment. PLoS Comput. Biol. 5:e1000392. doi:10.1371/journal.pcbi.1000392.
2. **Cai WZ, Schaffer PA.** 1989. Herpes simplex virus type 1 ICP0 plays a critical role in the de novo synthesis of infectious virus following transfection of viral DNA. J. Virol. 63:4579–4589.

2a. **Chowdhury S, Naderi M, Chouljenko VN, Walker JD, Kousoulas KG.** 2012. Amino acid differences in glycoproteins B (gB), C (gC), H (gH) and

Received 12 June 2012 Accepted 13 June 2012

Address correspondence to Stuart J. Macdonald, sjmac@ku.edu, or David J. Davido, ddavido@ku.edu.

doi:10.1128/JVI.01469-12

L (gL) are associated with enhanced herpes simplex virus type-1 (McKrae) entry via the paired immunoglobulin-like type-2 receptor. Virol. J. **9**:112.
3. **Cox MP, Peterson DA, Biggs PJ.** 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics **11**:485.
4. **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.
5. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**:R25.
6. **Otto TD, Dillon GP, Degrave WS, Berriman M.** 2011. RATT: rapid annotation transfer tool. Nucleic Acids Res. **39**:e57. doi:10.1093/nar/gkq1268.
7. **Perng GC, et al.** 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.
8. **Perng GC, et al.** 1995. An avirulent ICP34.5 deletion mutant of herpes simplex virus type 1 is capable of in vivo spontaneous reactivation. J. Virol. **69**:3033–3041.
9. **Roizman R, Knipe DM, Whitley RJ.** 2007. Herpes simplex viruses, p 2501–2601. *In* Knipe DM, et al (ed), Fields virology, 5th ed, vol 2. Lippincott Williams & Wilkins, New York, NY.
10. **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.
11. **Strelow LI, et al.** 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**(Part 9):2475–2480.
12. **Szpara ML, Parsons L, Enquist LW.** 2010. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. J. Virol. **84**:5303–5313.
13. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.