



# Molecular Evolution of Herpes Simplex Virus 2 Complete Genomes: Comparison between Primary and Recurrent Infections

Miguel A. Minaya,<sup>a</sup> Travis L. Jensen,<sup>b</sup> Johannes B. Goll,<sup>b</sup> Maria Korom,<sup>a\*</sup> Sree H. Datla,<sup>a</sup> Robert B. Belshe,<sup>c</sup> Lynda A. Morrison<sup>a,c</sup>

Department of Molecular Microbiology and Immunology, Saint Louis University School of Medicine, St. Louis, Missouri, USA<sup>a</sup>; The EMMES Corporation, Rockville, Maryland, USA<sup>b</sup>; Department of Internal Medicine, Saint Louis University School of Medicine, St. Louis, Missouri, USA<sup>c</sup>

**ABSTRACT** Herpes simplex virus 1 (HSV-1) and HSV-2 are large, double-stranded DNA viruses that cause lifelong persistent infections characterized by periods of quiescence and recurrent disease. How HSV evolves within an infected individual experiencing multiple episodes of recurrent disease over time is not known. We determined the genome sequences of viruses isolated from two subjects in the Herpevac Trial for Women who experienced primary HSV-2 genital disease and compared them with sequences of viruses isolated from the subsequent fifth or sixth episode of recurrent disease in the same individuals. Each of the HSV-2 genome sequences was initially obtained using next-generation sequencing and completed with Sanger sequencing. Polymorphisms over the entire genomes were mapped, and amino acid variants resulting from nonsynonymous changes were analyzed based on the secondary and tertiary structures of a previously crystallized protein. A phylogenetic reconstruction was used to assess relationships among the four HSV-2 samples, other North American sequences, and reference sequences. Little genetic drift was detected in viruses shed by the same subjects following repeated reactivation events, suggesting strong selective pressure on the viral genome to maintain sequence fidelity during reactivations from its latent state within an individual host. Our results also demonstrate that some primary HSV-2 isolates from North America more closely resemble the HG52 laboratory strain from Scotland than the low-passage-number clinical isolate SD90e from South Africa or laboratory strain 333. Thus, one of the sequences reported here would be a logical choice as a reference strain for inclusion in future studies of North American HSV-2 isolates.

**IMPORTANCE** The extent to which the HSV-2 genome evolves during multiple episodes of reactivation from its latent state within an infected individual is not known. We used next-generation sequencing techniques to determine whole-genome sequences of four viral samples from two subjects in the Herpevac Trial. The sequence of each subject's well-documented primary isolate was compared with the sequence of the isolate from their fifth or sixth episode of recurrent disease. Only 19 genetic polymorphisms unique to the primary or recurrent isolate were identified, 10 in subject A and 9 in subject B. These observations indicate remarkable genetic conservation between primary and recurrent episodes of HSV-2 infection and imply that strong selection pressures exist to maintain the fidelity of the viral genome during repeated reactivations from its latent state. The genome conservation observed also has implications for the potential success of a therapeutic vaccine.

**KEYWORDS** complete genome, Herpevac Trial for Women, evolutionary biology, genetic polymorphisms, herpes simplex virus, next-generation sequencing

Received 14 June 2017 Accepted 12 September 2017

Accepted manuscript posted online 20 September 2017

**Citation** Minaya MA, Jensen TL, Goll JB, Korom M, Datla SH, Belshe RB, Morrison LA. 2017. Molecular evolution of herpes simplex virus 2 complete genomes: comparison between primary and recurrent infections. *J Virol* 91:e00942-17. <https://doi.org/10.1128/JVI.00942-17>.

**Editor** Richard M. Longnecker, Northwestern University

**Copyright** © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Miguel A. Minaya, miguel.minaya@health.slu.edu, or Lynda A. Morrison, lynda.morrison@health.slu.edu.

\* Present address: Maria Korom, Department of Microbiology, Immunology and Tropical Diseases, George Washington University School of Medicine, Washington, DC, USA.

**H**erpes simplex virus 2 (HSV-2) is among the most successful human viruses in terms of its global distribution, evolutionary coexistence with humans, and persistence in the individual host (1–4). Current estimates suggest that HSV-2 afflicts more than 400 million people worldwide (5) and one-sixth of the U.S. population (6). HSV-2 primarily causes genital ulcerative disease (7). Babies are at risk of infection from their mothers during birth (8, 9). In addition, genital HSV infection increases the likelihood of HIV infection and transmission (10, 11). HSV-2 enters the body through the anogenital epithelium and then spreads via sensory nerve axons to the lumbosacral ganglia, where it establishes latent infection (12). Latent virus periodically reactivates due either to local stimuli, such as injury to tissues innervated by neurons harboring the virus, or to systemic stimuli, such as physical or emotional stress or hormonal imbalance (13). During a reactivation, the viral genome and virion components are synthesized in the nerve cell body and transported anterogradely within axons to the genital mucosa or skin at the original site of infection.

Periodic lesions caused by the reactivated virus (recurrent disease) or periods of asymptomatic shedding are the reservoirs for transmission from person to person (14, 15). Thus, limiting viral reactivation from the latent state is a central concern in prevention and treatment strategies. Common therapies for herpesvirus infection typically employ nucleoside analogs, such as acyclovir (ACV) and valacyclovir, which target the viral DNA polymerase. Although these drugs can be used episodically or chronically to treat HSV-2 infection (16, 17), they are incompletely effective (18, 19). In addition, resistance to the agents is an emerging problem for disease management (20–22). In an effort to protect HSV-naïve individuals, several recombinant vaccines to prevent HSV-2 genital infection have been clinically evaluated (23). Among them, a vaccine composed of glycoprotein D (gD) in alum and 3-O-deacylated monophosphoryl lipid A adjuvants showed promise in early vaccine trials (24), prompting a large, multicenter phase III trial in HSV-1/HSV-2-seronegative women, the Herpevac Trial for Women (25). The vaccine candidate provided 58% protection against HSV-1 disease but did not protect against HSV-2 infection or disease. Forty-four subjects developed primary HSV-2 disease, some of whom had multiple episodes of recurrent disease during the study period.

There is a paucity of information regarding factors that influence HSV-2 reactivation frequency, including the significance of individual isolates' genetic makeup. HSV-2 contains a linear, double-stranded DNA (dsDNA) genome of approximately 154,700 bp that has a GC content of 69% (26–28). At least 84 unique protein-coding open reading frames (ORFs) have been recognized, along with several RNA transcripts not proven to encode proteins (12). The HSV-2 genome consists of two covalently linked components, the unique long ( $U_L$ ) and unique short ( $U_S$ ) segments, which are flanked by inverted repeat long ( $IR_L$ ) and inverted repeat short ( $IR_S$ ) regions (29). Five genes are located within the  $IR_L$  and  $IR_S$  sequences and are therefore diploid (30). The large size and high GC content of the HSV-2 genome have been longstanding obstacles to understanding its molecular structure and evolution. The advent of next-generation sequencing, with its high throughput and low cost, has allowed investigation of the complete genomes of multiple geographically disparate HSV-2 strains (30, 31). Two of the completely sequenced and best-studied HSV-2 strains are HG52 and SD90e. HSV-2 HG52 is a high-passage-number laboratory strain from Scotland (27), while SD90e is a low-passage-number clinical isolate from South Africa (32). Compared to the HG52 strain, SD90e contains numerous single-nucleotide variants (SNVs), 13 insertions/deletions (indels), and 9 short compensatory frameshifts (33) that lead to differences in immunological and pathogenic properties (34). Because HG52 is less virulent than SD90e, it may not be representative of wild-type HSV-2 strains (34–37). It has therefore been suggested that the HSV-2 SD90e genome rather than that of HG52 should serve as the HSV-2 reference genome (30, 33).

Most investigations involving HSV genetics on the whole-genome scale have focused on characterization of drug resistance, geographic variation, phylogenetic associations, and recombination frequency (30, 31, 38). No studies to date compare the

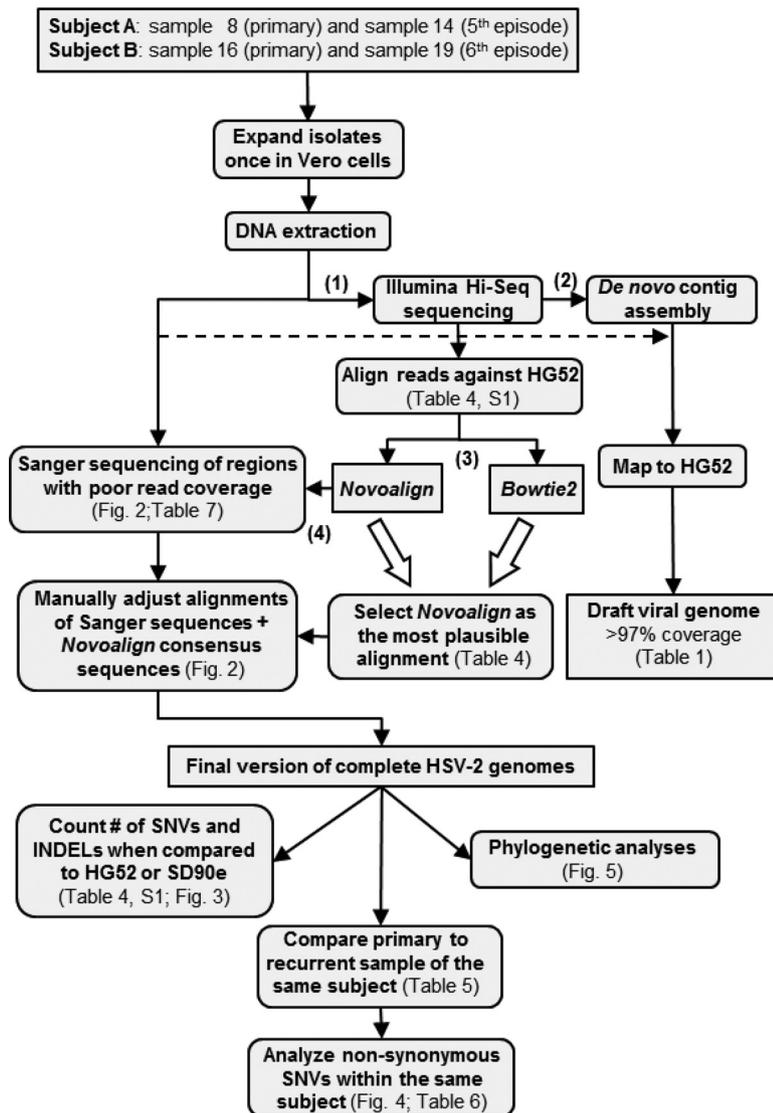
complete genome sequences of HSV-2 isolates from primary and recurrent episodes of disease. Understanding such viral genetic variability within an infected individual over time will facilitate research aimed at vaccine design, diagnosis, and therapeutic approaches. To begin to address this question, we used whole-genome sequencing to evaluate the frequencies and locations of nucleotide and amino acid changes in the HSV-2 genome that occur between primary infection and a recurrent disease episode in the same subjects.

## RESULTS

**Sequencing and genomic assembly.** Purified viral DNAs of primary HSV-2 isolates (samples 8 and 16) from two subjects in the Herpevac Trial and a later recurrent isolate (fifth or sixth symptomatic episode, samples 14 and 19) from the same subjects were sequenced using the Illumina platform (Fig. 1). The total number of reads for each viral isolate ranged from 2,568,857 to 29,742,520 (Table 1), the latter of which contained a large number of *de novo*-assembled contigs mapping to the African green monkey genome (contaminating cellular DNA). The viral genomic sequence coverage of the *de novo*-assembled contigs compared with the established HSV-2 strain HG52 genome ranged from 97.4% to 98% (Table 1). Per-base position qualities showed that, for all samples, the lower quartile of the quality score was  $>30$  for any base position (an incorrect base call probability of 1 in 1,000). While sample 8 had the lowest per-base position and average read quality compared to the other three samples, all Illumina reads passed the FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) default per-base and per-read sequence quality score checks. For all samples, the majority of reads were 45 bp or longer, with the primary and recurrent samples from subject A (samples 8 and 14, respectively) having a small percentage of reads around 25 bp in length (data not shown).

Reads were independently reference aligned using HG52 as the reference sequence. The number of mapped reads contributing to Novoalign (Novocraft) alignments ( $\sim 300,000$ ) and the number of plus- and minus-strand reads in Novoalign ( $\sim 121,000$ ) were similar across the four samples and lower than the number of mapped reads contributing to Bowtie2 (83) alignments. However, the Novoalign alignments resulted in up to a 3-fold-lower percentage of unique mapped reads than Bowtie2 (Table 2). Regions with low read coverage (read depth,  $<25$ ) were filled using Sanger sequencing (Fig. 1 and 2). The combination of the Novoalign sequence consensus and Sanger sequences yielded complete coverage of the HG52 HSV-2 genome for all four samples (Fig. 2), and they were uploaded to GenBank as a single contiguous sequence for each sample (Table 3).

Variant alleles detected in  $\geq 10\%$  of the read depth were tabulated with the ultimate goal of discerning genomic regions under evolutionary pressure. With HG52 used as the reference genome, the total number of SNVs in coding and noncoding positions within the  $U_S$  and  $U_L$  regions of all four samples combined was higher in the Bowtie2 alignments (560) than in the Novoalign alignments (485) (Table 4). Nonetheless, the percentages of total SNVs that fell within the ORFs were similar when the Novoalign (80.6%) and Bowtie2 (72.3%) alignments were compared, as were the percentages of nonsynonymous SNVs (41.4% and 39.8%, respectively). The total numbers of indels appearing in  $\geq 10\%$  of mapped reads in all four samples were also similar between the programmatic alignments (37 in Novoalign and 33 in Bowtie2), suggesting that neither of the programmatic alignments skewed the data set toward SNVs or indels (Table 4). One SNV resulting in a stop codon was found in a Novoalign alignment ( $U_L13$  gene; 14% of mapped reads; sample 16), and one more was found in the Bowtie2 alignments ( $U_L46$  gene; 11% of mapped reads; samples 8 and 14) (Tables 4 and 5). Based on this comparison, we consider the Novoalign alignment to be the most likely approximation for two reasons. First, the Novoalign alignments had fewer total SNVs in the four samples, and a smaller number of changes is the most plausible scenario according to evolutionary theories (39, 40). Second, the numbers of SNVs unique to the primary- or recurrent-disease samples in subjects A and B were 4- to 6-fold lower in the Novoalign



**FIG 1** Flow chart summarizing the data generation and analysis workflow. (1) HSV-2 was isolated from primary and recurrent infections in two subjects, and the viruses were expanded a single time in Vero cells. (2) Illumina Hi-Seq sequencing was performed on the whole genome of each sample. (3) *De novo* assembly of contigs was carried out using Velvet, and the contigs were aligned with the HG52 reference strain using Abacas/Nucmer to establish draft genome coverage rates. (4) Programmatic Novoalign and Bowtie2 alignments (BAM files), which excluded the IR<sub>L</sub> and IR<sub>S</sub> regions and used the HG52 strain as a reference, were compared based on the numbers of SNVs and indels that appeared in  $\geq 10\%$  of the Illumina read depth. Sanger sequencing was then performed on the HSV-2 regions with low read coverage (read depth,  $< 25$ ) based on Novoalign alignments. Sanger plus Novoalign consensus sequences for each sample (FASTA files) were aligned using nucleic and amino acid sequences. The HG52 and SD90e strains were used independently as reference genomes. Using the IUPAC code, SNVs observed in Novoalign alignments ( $\geq 10\%$  of the read depth) were incorporated into the manually adjusted alignment of each sample. The BAM and FASTA files were uploaded to GenBank (Table 3). The rectangular and rounded boxes represent, respectively, databases and actions done in the databases.

(subject A, 10; subject B, 9) than in the Bowtie2 (45 and 64, respectively) alignment (Table 4).

Using the consensus sequence of the Novoalign mapped reads plus Sanger sequences, the entire HSV-2 genomes obtained from the four samples were manually adjusted using, consecutively, the HG52 and SD90e reference sequences and then analyzed for polymorphisms. More SNVs and indels were found in coding and non-coding positions when we compared all four samples to the low-passage-number clinical isolate SD90e strain from South Africa (641 SNVs and 210 indels) than when we

**TABLE 1** *De novo* pseudomolecule assembly summary constructed using Velvet

Subject	Strain	No. of reads	Pseudomolecule length (bp)	No. of contigs (>100 bp)	No. of gaps in pseudomolecule	No. of overlapping contig separators in pseudomolecule	Genome sequence coverage (%)
A	Sample 8	3,088,997	174,355	577	182	144	97.7
	Sample 14	6,994,245	173,075	2,662	186	140	97.4
B	Sample 16	29,742,520	164,695	7,138	126	88	97.7
	Sample 19	2,568,857	170,115	436	141	108	98.0

compared them to the high-passage-number laboratory HG52 strain from Scotland (455 and 69, respectively) (Table 4). However, the percentages of SNVs that fell within ORFs were virtually the same regardless of whether the reference sequence used was HG52 (77%) or SD90e (65.8%). Nonsynonymous SNVs comprised similar percentages of the total SNVs whether the subjects' sequences were manually adjusted against the HG52 or the SD90e reference sequence (35.6% versus 27.3%, respectively). The stop codon found in a minority of mapped reads for the *U<sub>L</sub>13* gene (sample 16) remained in the alignment whether the reference strain was HG52 or SD90e. Overall, fewer SNVs and indels appeared in coding and noncoding positions in any of the four samples when HG52 was used as the reference (Table 4), and the numbers were very similar between isolates (data not shown).

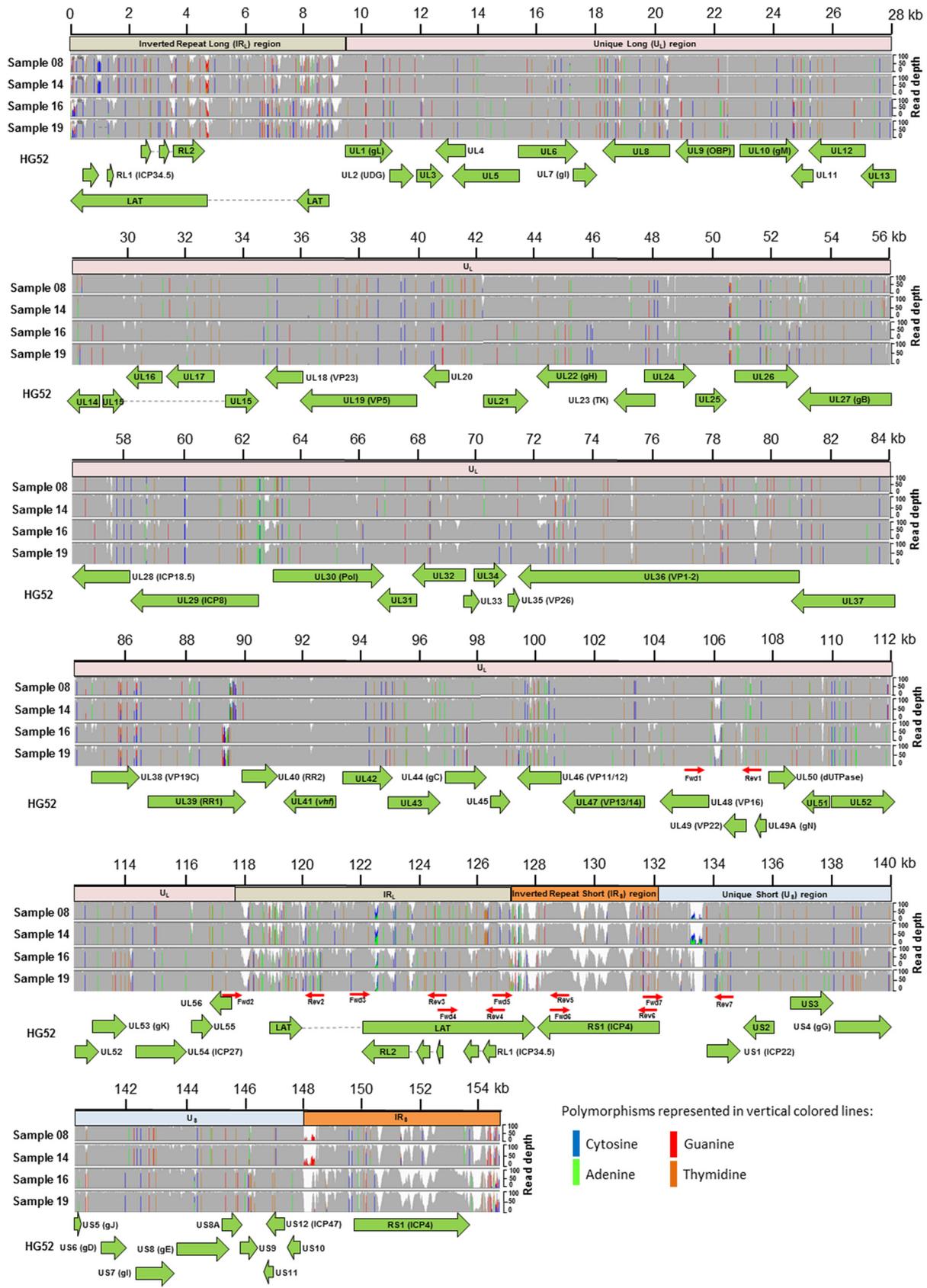
Using the manually adjusted alignments with the HG52 strain as the reference genome, we analyzed the sequence diversity of the individual ORFs. Our results indicate low sequence diversity and generally even SNV distribution across the coding regions (Fig. 3; see Table S1 in the supplemental material). The genes with the highest average percentages of SNVs were *U<sub>L</sub>39*, *U<sub>L</sub>49*, and *U<sub>S</sub>12*. *U<sub>L</sub>39* also had a high average frequency of nonsynonymous mutations, in addition to *U<sub>L</sub>3*, *U<sub>L</sub>11*, and *U<sub>L</sub>43*. No nucleotide variation was observed in *U<sub>L</sub>23*, *U<sub>L</sub>25*, *U<sub>L</sub>33*, *U<sub>L</sub>35*, *U<sub>L</sub>41*, *U<sub>L</sub>45*, *U<sub>L</sub>48*, *U<sub>L</sub>55*, *U<sub>S</sub>5*, *U<sub>S</sub>9*, or *U<sub>S</sub>11*.

**Unique polymorphisms within the primary or recurrent isolates.** Having assembled whole-genome sequences of primary and recurrent isolates from the same subjects provided the opportunity to examine how the HSV-2 genome evolves over time within an individual. Using the Novoalign alignments, we annotated unique polymorphisms in coding regions of each genome that represented  $\geq 10\%$  of the total reads at a given nucleotide position. A total of 10 nucleotide positions in the primary or recurrent isolate of subject A were altered in  $\geq 10\%$  of reads (Table 5). Seven of these nucleotide substitutions would produce a nonsynonymous change in the amino acid sequence of the gene. Similarly, 9 nucleotide positions varied in  $\geq 10\%$  of reads between the primary and recurrent isolates of subject B. Only five of them would alter the amino acid sequence of the encoded proteins. The unique polymorphic residue shared by subjects A and B was a synonymous mutation detected in the *U<sub>L</sub>14* gene (nucleotide position 321). In some cases, genetic variability at a nucleotide position within the primary-infection reads resolved to a single homogeneous nucleotide in the recurrent sample (e.g., nucleotide position 1039 in *U<sub>L</sub>21* of subject B); in other cases, more diversity occurred in the recurrent isolate (e.g., nucleotide 1133 in *U<sub>L</sub>13* of subject

**TABLE 2** Novoalign- and Bowtie2-based read-mapping statistics using HSV-2 strain HG52 as the reference genome

Subject	Strain	Isolate	No. of Novoalign/Bowtie2 reads				Mapped		Avg read depth
			Mapped	Unmapped	Plus strand	Minus strand	Unique <sup>a</sup> (%)	Nonunique	
A	Sample 8	Primary	302,552/1,033,661	2,786,445/2,055,336	122,049/368,973	121,982/363,735	7.90/23.72	58,521/300,953	96.17/209.68
	Sample 14	Recurrent (5th episode)	302,591/925,492	6,691,654/6,068,753	121,927/328,150	121,907/322,371	3.49/09.30	58,757/274,971	96.13/192.88
B	Sample 16	Primary	296,029/422,165	29,446,491/29,320,355	121,265/161,531	121,280/160,002	0.82/01.08	53,484/100,632	94.44/113.59
	Sample 19	Recurrent (6th episode)	299,862/334,781	2,268,995/2,234,076	121,754/131,109	121,799/130,814	9.48/10.20	56,309/72,858	95.69/100.69

<sup>a</sup>Uniquely mapped reads are defined as having a mapping quality (MQ) of  $\leq 30$ .



**FIG 2** Features of the complete HSV-2 genomes aligned with the HSV-2 strain HG52 reference sequence. Assembled Illumina mapped reads were aligned with the reference sequence and viewed in the IGV program. The scale at the top represents the genome position in kilobases. The horizontal bars below the genome scale represent the locations of the IR<sub>L</sub> regions (light brown), U<sub>L</sub> region (light pink), IR<sub>S</sub> regions (orange), U<sub>S</sub> region (light blue). (Continued on next page)

**TABLE 3** Genomes and accession numbers

Virus	Strain	Host	Isolation region	Primary clinical isolate	Collection yr	Country of origin	GenBank accession no. (FASTA/BAM files <sup>a</sup> )
HSV-1	17	Human	Unknown	No	1972	Scotland, United Kingdom	JN555585
ChHV-1	105640	Chimpanzee	Oral	Yes	2004	United States	NC023677
HSV-2	HG52	Human	Anal	No	Prior to 1971	Scotland, United Kingdom	JN561323
HSV-2	SD90e	Human	Genital	Yes	1994	South Africa	KF781518
HSV-2	8937_1999_3336	Human	Presumed genital	Yes	2003	United States	KR135298
HSV-2	10883_2001_13347	Human	Presumed genital	Yes	2005	United States	KR135311
HSV-2	9335_2005_576	Human	Presumed genital	Yes	2009	United States	KR135312
HSV-2	9335_2007_14	Human	Presumed genital	Yes	2011	United States	KR135313
HSV-2	7444_1996_25809	Human	Presumed genital	Yes	1996	United States	KR135314
HSV-2	89_390	Human	Presumed genital	Yes	1989	United States	KR135321
HSV-2	44_619833	Human	Presumed genital	Yes	2007	United States	KR135308
HSV-2	44_419851	Human	Presumed genital	Yes	2007	United States	KR135309
HSV-2	44_319857	Human	Presumed genital	Yes	2007	United States	KR135310
HSV-2	BethesdaP5	Human	Presumed genital	Yes	Unknown	United States	KR135330
HSV-2	333_R519	Human	Genital	No	1971	United States	KR135331
HSV-2	333	Human	Genital	No	1971	United States	KP192856
HSV-2	1192	Human	Genital	Yes	Unknown	United States	KP334095
HSV-2	COH 3818	Human	Presumed genital	Yes	Unknown	United States	KP334096
HSV-2	CtSF	Human	Presumed genital	Yes	Unknown	United States	KP334097
HSV-2	CtSF-R	Human	Presumed genital	Yes	Unknown	United States	KP334093
HSV-2	GSC-56	Human	Presumed genital	Yes	Unknown	United States	KP334094
HSV-2	Sample 8 <sup>b</sup>	Human	Genital	Yes	2004	United States	MF564034/MF783884
HSV-2	Sample 14 <sup>b</sup>	Human	Genital	Yes	2004	United States	MF564035/MF783885
HSV-2	Sample 16 <sup>b</sup>	Human	Genital	Yes	2007	United States	MF564036/MF783886
HSV-2	Sample 19 <sup>b</sup>	Human	Genital	Yes	2008	United States	MF564037/MF783887

<sup>a</sup>BAM files were obtained from the Novoalign mapped and unmapped reads. FASTA files were obtained from the final version of the whole HSV-2 genome (Fig. 1).

<sup>b</sup>Strain newly sequenced in this study.

B). Overall, these data demonstrate that the number of nucleotide positions undergoing change between the primary and a later recurrent isolate from the same individual was low.

It was of interest to determine whether nonsynonymous variants would produce a predicted change in protein conformation (Table 6). Among nonsynonymous variants in subject A, comparison of the predicted secondary and tertiary structures of the protein with those of a previously crystallized protein was only possible for U<sub>L</sub>13 (A190V and D229Y) and U<sub>L</sub>37 (H493P). Four nonsynonymous variants from subject B also could be examined by comparison to crystal structures: G378V and the minor variant R288STOP in U<sub>L</sub>13, A347T in U<sub>L</sub>21, and C951Y in U<sub>L</sub>30. A crystal structure of the G protein-coupled receptor kinase 2 (grk2) (41) (Protein Data Bank [PDB] accession no. 1YM7; residues 147 to 516) was used with 100% confidence to analyze the locations of the substitutions in the HSV-2 pUL13 protein kinase. Although residue 229 scored high for conservation and mutational sensitivity (Table 6), no change in the secondary  $\beta$ -sheet structure was observed (not shown). A crystal structure of the HSV-1 tegument protein pUL21 C-terminal domain (42) (PDB accession no. 5ED7; residues 277 to 527) was used with 100% confidence to evaluate the C-terminal domain of HSV-2 pUL21. The C-terminal domain of U<sub>L</sub>21 is composed of 10  $\alpha$ -helices and one  $\eta$ -helix arranged into a dragonfly fold with the left “wing” formed by  $\alpha$ 1 to  $\alpha$ 4 and  $\eta$ 1 (43, 44). Residue 347, located in helix  $\alpha$ 3 and present as threonine instead of alanine in 33% of the reads from subject B’s initial isolate, was strongly conserved and had a high score for

## FIG 2 Legend (Continued)

and U<sub>S</sub> region (light blue) in the genome of HSV-2 reference strain HG52. Samples are labeled on the left. BAM file reads were capped at 100 for visualization in IGV, and the read depth at each nucleotide position is represented in dark gray for each sample (obtained from IGV). The vertical colored lines inserted in the gray section represent polymorphisms in the genome using the HG52 strain as a reference. HSV-2 coding regions are shown as green arrows indicating the direction of the ORF. The red arrows represent the locations of forward and reverse primers used in Sanger sequencing (Table 7) to fill the regions of low read depth coverage.

**TABLE 4** Comparison of SNV and indel events among different alignments and reference genomes

Parameter	Value			
	Programmatic alignments <sup>a</sup>		Complete HSV-2 genome <sup>b</sup>	
	Novoalign	Bowtie2	HG52	SD90e
<b>SNVs</b>				
No. of SNVs in coding and noncoding positions (total of all four samples)	485	560	455	641
No. (%) of SNVs in ORFs (total of all four samples)	391 (80.6) <sup>c</sup>	405 (72.3)	350 (77)	422 (65.8)
No. (%) of SNVs in intergenic regions	94 (19.4)	155 (27.7)	100 (22)	191 (29.8)
No. (%) of SNVs in introns <sup>d</sup>	ND <sup>e</sup>	ND	5 (1)	28 (4.36)
No. (%) of nonsynonymous SNVs	201 (41.4)	223 (39.8)	162 (35.6)	175 (27.3)
No. of stop codons within an ORF creating possible pseudogenes	1 (U <sub>L</sub> 13)	1 (U <sub>L</sub> 46)	1 (U <sub>L</sub> 13)	1 (U <sub>L</sub> 13)
No. (%) of SNVs common to all four samples	191 (39.4)	216 (38.6)	149 (32.7)	214 (33.4)
No. (%) of SNVs only in subject A (samples 8 and 14)	129 (26.6)	180 (32.1)	146 (32)	164 (25.6)
No. (%) of SNVs only in subject B (samples 16 and 19)	165 (34)	164 (29.2)	160 (35.3)	164 (25.6)
No. of SNVs unique to the primary or recurrent infection in subject A	10	45	0	0
No. of SNVs unique to the primary or recurrent infection in subject B	9	64	0	0
<b>Indels</b>				
No. of indels in coding and noncoding positions (total of all four samples)	37	33	69	210
No. (%) of indels in ORFs (total of all four samples)	4 (10.81)	2 (6.06)	33 (47.8)	32 (15.2)
No. (%) of indels in intergenic regions	33 (89.2)	31 (93.9)	33 (47.8)	144 (68.6)
No. (%) of indels in introns <sup>d</sup>	ND	ND	3 (4.4)	34 (16.2)
No. (%) of indels common to all four samples	12 (32.4)	5 (15.1)	28 (40.6)	123 (58.6)
No. (%) of indels only in subject A (samples 8 and 14)	15 (40.5)	15 (45.4)	15 (21.7)	33 (15.7)
No. (%) of indels only in subject B (samples 16 and 19)	10 (27)	13 (39.4)	26 (37.7)	35 (16.6)
No. of indels unique to the primary or recurrent infection in subject A	5	0	0	0
No. of indels unique to the primary or recurrent infection in subject B	3	4	0	0

<sup>a</sup>The programmatic alignments, which included the high-quality mapped reads and HG52 strain as the reference sequence, were analyzed using SAMtools (v1.3) software. SNVs and indels were considered only when they appeared in  $\geq 10\%$  of the read depth.

<sup>b</sup>Complete HSV-2 genome obtained for each sample using Illumina and Sanger sequences improved manually using, consecutively, the nucleic acid and amino acid sequences. Only one copy of the IR<sub>L</sub> and IR<sub>S</sub> regions was included to compare the number of SNVs and indels observed when the HG52 or SD90e strain was used as a reference.

<sup>c</sup>The numbers in parentheses are the percentages of the total SNVs or indels.

<sup>d</sup>Programmatic alignments did not distinguish between intergenic regions and introns.

<sup>e</sup>ND, not determined.

mutational sensitivity (Table 6) but did not produce a change in the secondary  $\alpha$ -helix structure (not shown). A crystal structure of the HSV-1 DNA polymerase (45) (PDB accession no. 2GV9; residues 61 to 1202) was used with 100% confidence to model the HSV-2 pUL30 DNA polymerase. Four structural domains comprise pUL30. These domains assemble to form a disk-like shape around a central hole (45). The C951Y substitution found in 20% of the reads from subject B's primary isolate (Table 6) produced an extension of the  $\beta$ -sheet located within the palm subdomain of the polymerase domain but did not otherwise alter its structure (not shown). The P1133L substitution in 60% of subject A's recurrent isolate could not be analyzed because amino acids (aa) 1100 to 1138 were not included in the protein that had been crystallized. Although subjects A and B received valacyclovir therapy for their infections, we noted no decrease in ACV sensitivity of their recurrent isolates (data not shown), consistent with the genetic sequence information. A crystal structure of the pseudorabiesvirus pUL37 N-terminal half (46) (PDB accession no. 4K70; residues 24 to 570) was used with 100% confidence to evaluate the secondary and tertiary structures of HSV-2 pUL37. Domain III of pUL37 is a highly conserved helical bundle, with a central helix ( $\alpha$ 19) surrounded by six helices ( $\alpha$ 16 to  $\alpha$ 22) (46). Ninety-three percent of the reads from subject A's primary isolate specified proline at residue 493 (Table 6), whose replacement by histidine in the recurrent isolate disrupted helix  $\alpha$ 21 within domain III of pUL37 (Fig. 4).

**Phylogenetic reconstruction.** Phylogenetic associations among the HSV-2 primary- and recurrent-disease samples; previously published unverified (30) or partial (31) HSV-2 genome sequences from North America; and the HSV-2 reference strains HG52 (Scotland), SD90e (South Africa), and 333 (North America) were evaluated using HSV-1

and chimpanzee herpesvirus 1 (ChHV-1) genomes as outliers. The phylogenetic reconstruction based on the whole HSV-2 genome provides a graphical representation of the relationships among the samples (Fig. 5). Expansion of the HSV-2-specific nodes (Fig. 5B) indicated a close relationship between the isolates from subject A (samples 8 and 14) and those from subject B (samples 16 and 19). It also suggested a close relationship between the laboratory HG52 strain from Scotland and the samples from subject A, while the low-passage-number clinical isolate SD90e from South Africa appeared to be strongly nested within a group that includes other North American strains published previously.

## DISCUSSION

This research presents for the first time a comparison of whole HSV-2 genome sequences of primary isolates and isolates from later recurrent-disease episodes from two infected individuals. Coverage of >97% was reached based on *de novo* pseudo-molecule assembly contig mapping compared to the HSV-2 HG52 strain, and 100% coverage was reached when we combined Novoalign sequence alignments and Sanger sequences (Fig. 2). The two individuals became infected while they participated in the Herpevac Trial for Women (25). Two of the genome sequences derive from the subjects' primary isolates and two of them from the well-documented fifth (subject A) or sixth (subject B) recurrent episodes of genital disease. Although various mechanisms of natural variation were evident, including SNVs and indels, our analyses demonstrate a high degree of sequence homology when each subject's primary and recurrent isolates were compared, suggesting a low rate of evolution over time in an infected individual. The striking sequence conservation between the primary isolate and a later recurrent isolate from the same subjects suggests strong selection pressures on the virus to maintain genetic fidelity during reactivations from its latent state, but more temporally distant isolates will be needed to firmly establish the rate of genetic drift or the impact of immune selection.

Identification of unique polymorphisms in primary infection versus recurrent episodes of disease throughout the complete HSV-2 genome sequence may help us to ascertain regions of the genome undergoing molecular evolution during reactivation events. The strong sequence conservation we observed, whether in a subject with a high frequency of recurrence (subject A) or less than once per month (subject B), also augers well for the success of a therapeutic vaccine because it suggests evolution of the viral genome is constrained by selective pressures, including the immune response in an infected individual. The apparent rate of mutational drift in HSV is much lower than in other persistent virus infections where it has been studied, such as HIV and hepatitis C virus (47). Possible reasons include the proofreading activity of the HSV DNA polymerase and HSV's complex life cycle that must accommodate the restrictions imposed by diverse necessities of existing in a latent state in neurons and rapidly replicating in mucosal epithelia.

All Illumina reads passed the FastQC default per-base and per-read sequence quality checks. The Illumina coverage distribution observed in our results was similar to that previously seen for HSV-1 (48) and HSV-2 (31), with lower coverage in the repeat regions and higher coverage in the  $U_L$  and  $U_S$  coding regions. The ambiguous regions of Illumina sequence were filled using conventional Sanger sequencing, with special conditions for GC-rich DNA, as has been previously published for sequencing of other strains of HSV-2 (33) and HSV-1 (38). The combination of the assembled Illumina reads and Sanger sequences yielded a full-length version of the HSV-2 genomes for samples 8, 14, 16, and 19 (Fig. 2).

Phylogenetic analyses were performed to further evaluate the association among the primary and recurrent disease isolates, other North American HSV-2 strains previously published, and the HG52 and SD90e strains, using HSV-1 strain 17 and ChHV-1 strain genomes as outgroups (Fig. 5). In contrast to previous studies that focused on selected regions of the genome or excluded intergenic sequences (49, 50), we used the full genome alignment, including coding and noncoding positions, to assess relation-

**TABLE 5** SNVs unique to the primary or recurrent infection in subject A or B

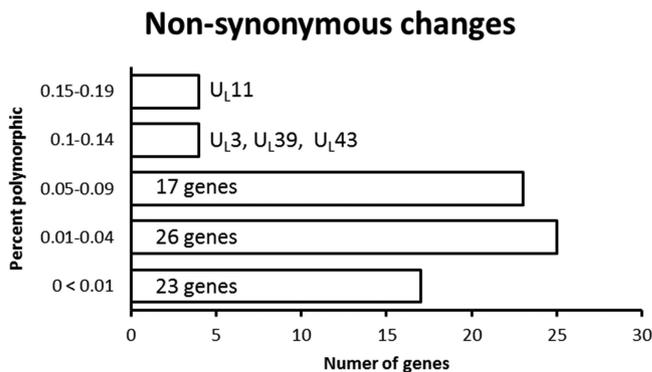
Gene	Nucleotide variant		Amino acid variant		Function	
	Nucleotide/ amino acid position	Primary infection (sample 8) >> recurrent infection (sample 14)	Primary (sample 16) >> recurrent (sample 19)	Primary infection		Recurrent infection
Subject A						
U <sub>L</sub> 13 <sup>a,b</sup>	569/190	C (86) <sub>c</sub> , T (14) >> C (98), T (1), A (1)		A (86), V (14)	A (98)	Protein kinase
U <sub>L</sub> 13 <sup>b</sup>	685/229	G (100) >> G (90), T (10)		D (100)	D (90), Y (10)	Protein kinase Tegument protein
U <sub>L</sub> 14	321/107	T (10), C (90) >> T (74), C (26)				Tegument protein
U <sub>L</sub> 14	334/112	C (70), T (30) >> C (99), A (1)		R (70), C (30)	R (100)	Tegument protein B
U <sub>L</sub> 27	206/69	C (100) >> T (52), C (48)		P (100)	L (52), P (48)	Glycoprotein B
U <sub>L</sub> 30	3398/1133	C (98), T (1), A (1) >> T (60), C (40)		P (98)	L (60), P (40)	DNA polymerase
U <sub>L</sub> 36	7942/2648	G (100) >> G (65), A (35)		A (100)	A (65), T (35)	Tegument protein
U <sub>L</sub> 37 <sup>b</sup>	1478/493	A (7), C (93) >> A (99), G (1)		P (93)	H (99)	Tegument protein
U <sub>L</sub> 44	321/107	G (82), C (15), A (1), T (2) >> G (91), C (8), T (1)				Tegument protein
Intergenic space	145807 <sup>d</sup>	G (89), A (11) >> G (99), A (1)				
Subject B						
U <sub>L</sub> 12	1305/435		C (100) >> T (24), C (76)			DNase
U <sub>L</sub> 13 <sup>b</sup>	862/288		C (86), T (14) >> C (98), T (2)	R (86), STOP (14)	R (98)	Protein kinase
U <sub>L</sub> 13 <sup>b</sup>	1133/378		G (100) >> G (86), T (14)	G (100)	G (86), V (14)	Protein kinase
U <sub>L</sub> 14	241/81		G (99), T (1) >> G (79), A (21)	V (99)	V (79), M (21)	Tegument protein
U <sub>L</sub> 14	321/107		T (81), C (19) >> T (45), C (55)			Tegument protein
U <sub>L</sub> 14	396/132		C (100) >> C (90), T (10)			Tegument protein
U <sub>L</sub> 21 <sup>b</sup>	1039/347		G (67), A (33) >> G (100)	A (67), T (33)	A (100)	Tegument protein
U <sub>L</sub> 24	165/55		G (89), A (11) >> G (100)			Nuclear protein
U <sub>L</sub> 30 <sup>b</sup>	2852/951		G (80), A (20) >> G (100)	C (80), Y (20)	C (100)	DNA polymerase

<sup>a</sup>Boldface indicates nonsynonymous substitutions.

<sup>b</sup>The secondary and tertiary structures of the protein that contained the polymorphisms were analyzed using the Phyre2 Web portal (Table 6).

<sup>c</sup>The numbers in parentheses are the percentages of Novoalign mapped reads supporting each polymorphism. The reported polymorphisms appeared in  $\geq 10\%$  of mapped reads. Underlining indicates the polymorphic variant allele.

<sup>d</sup>Polymorphism position observed in an intergenic space derived from the Novoalign alignments (BAM files).



**FIG 3** Number of HSV-2 genes at each interval of nonsynonymous change. The intervals (y axis) were selected based on the average percentage of SNVs that represented nonsynonymous changes across ORFs, using the HG52 strain as the reference sequence. The most polymorphic genes are identified. See Table S1 in the supplemental material for a list of all the SNVs observed in the ORFs of the complete HSV-2 genomes.

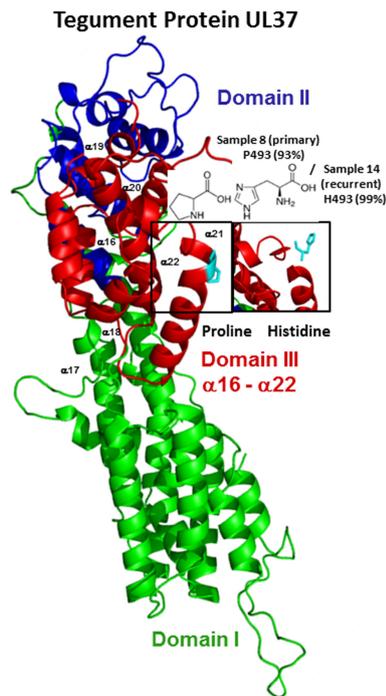
ships among strains. Several salient observations arose. First, our results demonstrate that the primary isolates (samples 8 and 16) cluster with their respective recurrent-disease isolates (samples 14 and 19) rather than with each other (Fig. 5B). Therefore, accumulated SNVs that occur with repeated reactivations of HSV-2 from a latent state do not necessarily lead to sequence convergence when considering the genome as a whole. Second, there is a strongly supported cluster between the HG52 strain and the samples from subject A (samples 8 and 14) and subject B (samples 16 and 19), while the SD90e and 333 strains appear to be nested within groups that include other North American strains (30, 31). Thus, comparison of the sequences from subjects A and B with genomes from other North American HSV-2 isolates indicates that our sequences represent a distinct group with the greatest homology to HG52 (Fig. 5B). These phylogenetic observations corroborate our numerical observations: our four HSV-2 genome sequences from North American clinical isolates (subjects A and B) had fewer SNVs and indels than the high-passage-number laboratory HG52 strain ([JN561323.2](#)) from Scotland and the low-passage-number clinical isolate SD90e ([KF781518.1](#)) from South Africa (Table 4). Our results also agree with the low sequence diversity previously observed among HSV-2 genomes (30, 33). Thus, although the HSV-2 SD90e strain is a low-passage-number clinical isolate, our observations suggest a closer association between some North American samples and the Scotland strain HG52 than most other North American isolates previously reported (30, 33). Based on this result, one of the HSV-2 complete genomes of clinical isolates reported here would be a logical choice as a reference strain for inclusion in future European and North American studies.

Using the HG52 strain as a reference genome, we analyzed the complete HSV-2 genomes for hot spots in the polymorphism distribution and also ORFs with the highest number of SNVs and nonsynonymous changes, suggestive of proteins that are under-

**TABLE 6** Phyre2 investigator tests

Gene	Subject	Sample	Isolate	Amino acid position	Amino acid variant (%)	Secondary structure	Structural modifications	ProQ2 quality	Alignment confidence	Residue conservation	Mutational sensitivity
U <sub>L</sub> 13	A	8	Primary	190	A (86), V (14)	α-Helix	No	Medium	Medium	Low	Low
U <sub>L</sub> 13	A	14	Recurrent	229	D (90), Y (10)	β-Sheet	No	Medium	Good	High	High
U <sub>L</sub> 13	B	16	Primary	288	R (86), stop (14)	Connector					
U <sub>L</sub> 13	B	19	Recurrent	378	G (86), V (14)	Connector	No	Medium	Good	Low	Low
U <sub>L</sub> 21	B	16	Primary	347	A (67), T (33)	α-Helix	No	Good	Good	High	High
U <sub>L</sub> 30	B	16	Primary	951	C (80), Y (20)	β-Sheet	No	Good	Good	Low	Low
U <sub>L</sub> 37 <sup>a</sup>	A	8	Primary	493	P (93)	α-Helix	Yes	Medium	Good	Low	Low
U <sub>L</sub> 37 <sup>a</sup>	A	14	Recurrent	493	H (99)						

<sup>a</sup>Represented in Fig. 4.



**FIG 4** Features of the amino acid variants resulting from nonsynonymous changes in the primary or recurrent isolates of the tegument protein  $U_L37$ . Assessment of the tertiary molecular structure of pUL37 is shown (with structural modifications) based on the crystal structure of the N-terminal half of pseudorabiesvirus pUL37 (PDB accession no. 4K70; residues 24 to 570). Domains are numbered and represented in different colors. The polymorphic amino acids (P493H) (Table 6) are boxed and colored in cyan on the protein tertiary structures.

going evolution. Such regions of higher sequence lability may allow a variant to emerge that could, for example, evade host immune surveillance or adapt to a new host's genetic makeup. The regions with the highest levels of variation appeared in the introns and intergenic sequences (Fig. 2; see Table S2 in the supplemental material), consistent with recently published nearly complete HSV-2 genomes (30, 33). Numerous ORFs contained SNVs relative to the reference sequence, but these nucleotides did not differ between primary- and recurrent-disease isolates. Therefore, the primary-infection sample may have fixed a polymorphism that confers a selective advantage or simply may have tolerated diversity at a given location with little or no selective pressure. Unique SNVs in ORFs of the primary or recurrent HSV-2 samples were primarily concentrated in two genes,  $U_L13$  and  $U_L14$  (Table 5).  $U_L13$ , which encodes a protein kinase, contained four unique nonsynonymous polymorphic residues. Two of the polymorphic nucleotides specified changes in samples from the primary infection (samples 8 and 16) that were subsequently purified to near homogeneity in the recurrent-disease samples; the other two represent diversifying mutations in the recurrent samples (samples 16 and 19).  $U_L14$ , which encodes a tegument protein, contained five unique SNVs. Two of these polymorphisms were nonsynonymous, one of which generated more diversity in the recurrent sample (sample 19) while the other resulted in near homogeneity in the recurrent sample (sample 14). Interestingly, 4 out of 5 unique SNVs in  $U_L14$  resulted in greater diversity in the recurrent sample (Table 5); the  $U_L14$  gene also showed a high degree of sequence variability in previous studies of HSV-1 (38) and HSV-2 (27). The read depth heterogeneity at all 19 of the SNVs among the four samples was  $\geq 10\%$ , lending credence to the interpretation that these were true polymorphisms and not sequencing errors. Though our sample size was small, the low number of unique SNVs suggests strong conservation of the genome sequences between the primary and recurrent HSV-2 isolates from the same subject. While this result implies substantial selection pressure on the viral genome during its history



detailed road map and an excellent opportunity to explore the structural/phenotypic implications of the amino acid changes unveiled. Modeling of the polymorphisms in pUL13, pUL21, and pUL30 did not reveal an appreciable change in structure (data not shown), despite prediction of two polymorphisms as mutationally sensitive for proper structure and function of the protein according to the Jensen-Shannon divergence (51) and the SuSPect method (52). pUL13 is a Ser/Thr protein kinase conserved among alpha-, beta-, and gammaherpesviruses (53) and has important roles in the herpesvirus replication cycle (54, 55), tegument dissociation (56), viral-gene expression (57, 58) and cell cycle regulation (59). Despite its many important roles, four genetic polymorphisms unique to the primary or recurrent samples were found in subject A or B. However, none of the four polymorphisms identified fell within a known kinase domain (60). Unlike HSV-1 U<sub>L</sub>13 mutants, HSV-2 U<sub>L</sub>13-deficient mutants are highly compromised for replication (57, 61, 62; M. Korom, J. E. Schrimpf, G. S. Delassus, and L. A. Morrison, unpublished data). The stop codon predicted at pUL13 residue 288 in 14% of the reads in primary-infection sample 16 would therefore decrease viral fitness, so it is not surprising that this polymorphism disappeared in the recurrent-infection sample 19. pUL37 is essential for HSV-1 replication, plays an important role in capsid trafficking during virion entry, and interacts with the gK-UL20 protein complex to facilitate cytoplasmic virion envelopment (63–67). The P493H substitution in helix  $\alpha$ 21 lies in the highly conserved helical bundle domain III of pUL37 and is predicted to disorder a portion of the helix (Fig. 4). Residue 493 is close to a pUL37 self-association domain (aa 568 to 1123), which is one of its domains that appear to have distinct functions during virus replication (67). Thus, the polymorphism observed in residue 493 may have functional consequences, although the conservation and mutational sensitivity of the residue are predicted to be low. Nonetheless, it is of interest that the proline that predominated at this residue in the primary sample and is present in herpesviruses of several animal species (68) had been replaced in 99% of the reads of the recurrent sample. Its proximity to tyrosine 480, which interacts with gK to facilitate cytoplasmic virion envelopment and infectious-virus production (67), suggests that pressure to revert to histidine may exist to maximize capsid egress and envelopment.

This research presents for the first time a comparison of HSV-2 complete genome sequences representing temporally distinct isolates from the same subjects based on next-generation and Sanger sequencing techniques. Each subject's primary isolate and later recurrent isolate were highly homologous. The differences described among the four low-passage-number HSV-2 strains newly sequenced in this research, the high-passage-number HG52 strain, and the low-passage-number SD90e strain support the use of one of the sequences reported here, derived from primary clinical isolates from the United States, as a reference strain for future European and North American studies because they represent a subset of sequences with greater similarity to HG52 than most other sequences or partial sequences of North American isolates determined to date (30, 31). Although we have been able to answer initial questions about the amount of diversity among temporally distinct HSV-2 isolates from a single individual, and our data provide a context for future evolutionary and virulence studies, further investigations of genome diversity among primary HSV isolates will be important to development of a broadly effective prophylactic or therapeutic vaccine.

## MATERIALS AND METHODS

**Strain sampling, virus DNA isolation, and sequencing.** HSV-2 sequences were generated from DNA extracted from primary and recurrent isolates from two healthy, initially HSV-1/HSV-2-seronegative subjects (A and B) collected during the Herpevac Trial for Women (25). The times of initial infection and recurrent disease were well documented because the subjects became infected while participating in the Herpevac Trial. The viral isolates were designated sample 8 (primary isolate, subject A), sample 14 (recurrent isolate, fifth episode, subject A), sample 16 (primary isolate, subject B), and sample 19 (recurrent isolate, sixth episode, subject B) (Table 2 and Fig. 1). Subject A received HSV-2 gD vaccine and became infected approximately 8 weeks after the final dose; she experienced recurrences an average of every 20 days. Subject B received control vaccine, and her infection occurred approximately 6 weeks after the final dose; she experienced recurrences an average of every 44 days. The subjects provided written consent to future use of their samples, and sequencing of the isolates was approved by the Saint Louis

**TABLE 7** HSV-2 amplification and sequencing primers

Primer <sup>a</sup>	Sequence	Annealing temp (°C)	Genome position
Fwd1_HSV-2	5'-GTT GGG TCC GGG AAT AAC GA-3'	58	U <sub>L</sub> 48 (part), intergenic region, U <sub>L</sub> 49 (part)
Rev1_HSV-2	5'-CGC GAA CGA GTT GGT GAA TC-3'	58	U <sub>L</sub> 48 (part), intergenic region, U <sub>L</sub> 49 (part)
Fwd2_HSV-2	5'-ATT AAC GCA CGC ATG CAG AC-3'	58	U <sub>L</sub> 56 (part), intergenic region
Rev2_HSV-2	5'-CTT CTG CCC TTC CAT CCT CC-3'	58	U <sub>L</sub> 56 (part), intergenic region
Fwd3_HSV-2	5'-GTT TCT CGT CTC TCC CCA GC-3'	55	LAT (exon 2), RL2 (part)
Rev3_HSV-2	5'-GTG GCC TCT CTT CCC CCT-3'	55	LAT (exon 2), RL2 (part)
Fwd4_HSV-2	5'-CTG CAG AGG GAG ACA GAG A-3'	62	LAT (exon 2), RL1 (part)
Rev4_HSV-2	5'-GGG AAT CTC TGA CGA CGA CC-3'	62	LAT (exon 2), RL1 (part)
Fwd5_HSV-2	5'-TCT CCC AGG CCA CCA GAT G-3'	64	LAT (exon 2), RS1 (part)
Rev5_HSV-2	5'-TGA GTT CGC TAG GCA AGC AC-3'	64	LAT (Exon2), RS1 (part)
Fwd6_HSV-2	5'-ACG CAG GGA CCA TTT GGG AGT C-3'	55	RS1
Rev6_HSV-2	5'-CCT GAG TGC AGG TTA CG-3'	55	RS1
Fwd7_HSV-2	5'-GCG CAT CGG TTC CTT TTC G-3'	58	Intergenic region, U <sub>S</sub> 1 (part)
Rev7_HSV-2	5'-GTC GGG CTT ACC CTC AGA TT-3'	58	Intergenic region, U <sub>S</sub> 1 (part)

<sup>a</sup>Used for amplification and sequencing.

University Institutional Review Board (IRB number 24706). Virus was isolated in Vero (African green monkey kidney) cell monolayers directly from the thawed clinical swabs in transport medium. The Vero cells were originally acquired from the laboratory of David Knipe. The isolates were then inoculated into larger flasks of Vero cells at a multiplicity of infection (MOI) of 0.1, and supernatant and cells were collected when the cytopathic effect reached 100%. Fresh DNA stocks were prepared by cesium chloride gradient sedimentation as previously described (69). Purified genomic DNA (500 ng) was submitted to the Genome Technology Access Center (St. Louis, MO). Sequence libraries were constructed by bar code addition to sheared DNA and run on a single lane on an Illumina Hi-Seq sequencing machine, producing 50-bp single-end reads. The quality and quantity of the DNA were assessed using an Agilent DNA high-sensitivity series chip assay (Agilent Technologies, Santa Clara, CA, USA) and a Qubit dsDNA kit (Life Technologies, Grand Island, NY, USA), respectively, and the libraries were standardized to 2 nM. The sequencing lane contained a PhiX plasmid spike-in; 0.6% of the reads were PhiX and demonstrated an error rate of 0.2% (1 per 500 bases).

**HSV-2 genome sequence alignments using a reference-based approach.** (i) **De novo pseudomolecule assembly.** The Velvet software suite (v1.2.08) (84) was utilized for *de novo* contig assembly of all Illumina reads available for each sample (Table 1 and Fig. 1). Maximum K-mer length was set to 31, and the minimum contig length was set to 100. To obtain a genome scaffold (reference genome-oriented best-matching contigs), the Abacas (v1.3.1)/Nucmer software package (85, 86) was used applying the following consecutive steps. (Step 1) Repeat regions were handled via stepwise masking. The reference genome was masked in three separate locations in order to best align contigs to all repeat regions. The first unmasked region ranged from the beginning of the viral genome through the U<sub>L</sub> region and up to the nucleotide base before the IR<sub>L</sub> region. The second unmasked region ranged from the first base of the IR<sub>L</sub> region through the IR<sub>S</sub> region and up to the nucleotide base before the U<sub>S</sub> region. The final unmasked region covered the first nucleotide base of the U<sub>S</sub> region through the end of the genome. (Step 2) Contig mapping results for each partially masked reference genome comparison were combined and assembled into a contiguous pseudomolecule or DNA sequence assembly. Partially overlapping contigs were separated by 100 Ns. The base and read qualities of sequencing reads were assessed using the FastQC tool kit (v0.10.1). Finally, comparison of the *de novo*-assembled pseudomolecules with the HSV-2 HG52 genome using Nucmer indicated more than 97% coverage was achieved for each molecule (Table 1).

(ii) **Programmatic alignments: Novoalign and Bowtie2.** The Illumina reads were independently aligned with the sequence of the high-passage-number clinical isolate HG52 laboratory strain (from Scotland) and analyzed using Novoalign (v3.02.06) and Bowtie2 (v2.1.0) mapping software (Table 2 and Fig. 1). For Novoalign, reads were selected for alignment if (i) 30 or more bases were of good quality, (ii) repetitive DNA sequences (homopolymers) had a score of >90, and (iii) dinucleotide sequences had a score of >120. The maximal alignment score was set to 254, a limit of 100 was set for the number of times a single read could be aligned, and repeats were reported in a random fashion. Bowtie2-based reference alignments utilized the -fast flag (implies -D 10 -R 2 -N 0 -L 22 -i S,0,2,50). Alignment quality control (QC) statistics for each binary alignment/map (BAM) file were obtained using RSeQC software (v2.3.7) (87). The resulting Novoalign BAM files were visualized against the HG52 reference genome using the Integrative Genomics Viewer (IGV) program (70) with a cap of 100 reads.

(iii) **Sanger sequencing of regions with low read depth.** Genome regions with low coverage in Novoalign reference mappings or incomplete/improper overlap in *de novo* assembly were filled using the Sanger sequencing method (Fig. 1). HSV-2 sequences were PCR amplified from 100 ng of the cesium chloride gradient-purified viral DNA using the primers listed in Table 7. Amplification reactions were performed using a reaction mixture containing 0.75  $\mu$ l of forward and reverse primers (10 mM), 2.5  $\mu$ l 10 $\times$  AccuPrime Pfx reaction mix (Invitrogen), 2.5  $\mu$ l betaine solution (5 M) (Sigma), 2  $\mu$ l MgCl<sub>2</sub> (50 mM), 1.5  $\mu$ l dimethyl sulfoxide (DMSO), 0.5  $\mu$ l Taq DNA polymerase, 2 to 4  $\mu$ l of template DNA, and sterile ultrapure water (MilliQ) added to achieve a final volume of 25  $\mu$ l. The amplification parameters consisted of a primary denaturing step of 2 min at 95°C, followed by 39 cycles of 20 s denaturing at 95°C, 30 s

annealing, and 3 min extension at 68°C, followed by a final extension step of 5 min at 68°C. The PCR products were run on agarose gels to confirm the size and purity of the amplicon. The PCR products were then cleaned up using ExoSap-It (Affymetrix) following the protocol indicated by the manufacturer and sent to GeneWiz, Inc. (South Plainfield, NJ) for Sanger sequencing using their GC-rich sequencing and difficult-template protocols. Complementary strands were assembled and verified using Sequencher version 4.2.2 (Gene Codes Corp.).

**(iv) Manual alignments.** The consensus sequence of the Novoalign mapped reads was extracted and saved as a FASTA file for each of the four samples (Fig. 1). The four FASTA files were imported into MEGA (v7.0.14) (71) software, along with sequences of the HG52 strain (from Scotland) and the low-passage-number clinical isolate SD90e (from South Africa), which were used as reference sequences. The Sanger sequences were then incorporated into the FASTA file of each sample. The genome sequence alignments obtained for each sample were improved manually using, consecutively, the nucleic acid and amino acid sequences. The full-length version of each sample genome was created by placing inverted copies of the IR<sub>L</sub> and IR<sub>S</sub> regions at the appropriate termini. The consensus sequence for each genome was generated by designating any variant of >90% prevalence as the consensus base. Variant alleles detected in ≥10% of the read depth were annotated in the consensus sequence using the International Union of Pure and Applied Chemistry (IUPAC) nucleotide code. Variants with no clear majority allele (those with <10% prevalence) were omitted in the consensus sequence.

**Evolutionary analyses. (i) Evaluation of SNVs and indels.** To select the most plausible alignment according to evolutionary theories (39, 40), SNVs and indels were tabulated based on the Novoalign and Bowtie2 programmatic alignments. The programmatic alignments, which included the high-quality Illumina reads and used the HG52 strain as a reference, were analyzed using SAMtools (v1.3) software (88). SAMtools mpileup was used on BAM-formatted mapped reads to generate information on match, mismatch, indel, strand, and mapping quality per reference genomic position. To reduce the impact of false positives on variant calling, only high-quality mapped reads and reads with base quality and mapping quality of ≥20 were utilized in the pile-up. SNVs and indels were also tabulated from the consensus sequence for each genome generated after manual comparison to HG52 and independently to SD90e as the reference strains, using the MEGA (v7.0.14) software.

**(ii) Analysis of unique SNVs and modeling of protein structural changes.** Some SNVs represented in ≥10% of the read depth appeared exclusively in the primary or recurrent sample from the same subject. Among these, amino acid variants resulting from nonsynonymous changes were mapped and analyzed based on the secondary and tertiary structures of a previously crystallized protein homolog with confidence of >99%. These comparisons were performed in the Phyre2 (Protein Homology/analogy Recognition Engine V2.0) Web portal (71). The Phyre investigator tool was used to analyze the features of the amino acid sequences of the HSV-2 proteins compared. The analyses included (i) prediction of the secondary structure of the HSV-2 protein of interest and structural modifications thereto by comparison to a crystallized protein homolog, (ii) assessment of the quality of the HSV-2 protein model predicted using ProQ2 (72), (iii) analysis of the consistency of the pairwise query template alignment (alignment confidence) based on the posterior probabilities calculated in the forward-backward algorithm (these values were calculated by Phyre investigator, scanning our sequence against a sequence database using the iterative sequencing program PSI-Blast), (iv) prediction of whether an altered amino acid residue was essential for proper structure and function of the protein (residue conservation) based on the Jensen-Shannon divergence (51), and (v) estimation of the effects of mutations at a particular position in our sequence (mutational sensitivity) using the SuSPect method (52). Lastly, the PyMOL Molecular Graphics System, version 1.8 (Schrödinger LLC) was used for rendering and editing the secondary and tertiary molecular structures found by the Phyre2 Web portal.

**(iii) Phylogenetic analyses.** A phylogenetic reconstruction was used to assess the evolutionary relationships among our four samples, other North American strains published previously (30, 31), and the reference HSV-2 strains HG52 and SD90e. The HSV-1 strain 17 and ChHV-1 genome sequences were aligned and used as outgroups against the HSV-2 manually adjusted data set (Table 3). The final version of the whole genome of samples 8, 14, 16, and 19 was used, which included all SNV and indel events (≥10% of the read depth) marked using the IUPAC code (Fig. 1). Bayesian inference (BI) analyses and maximum-likelihood (ML)- and maximum-parsimony (MP)-based searches (including coding and non-coding positions) were performed using, respectively, MrBayes 3.2.6 (73, 74), RAxML 7.2.8, and PAUP\* 4.0 beta 10 (75). Tests of goodness of fit for alternative nucleotide substitution models were performed through the Akaike information criterion (AIC) (76) and Bayesian information criterion (BIC) (77) tests and a decision-theoretic (DT) performance-based approach (78) in jModelTest 2 (79, 80). jModelTest 2 selected TVM+G as the optimal model. The closest GTR+G model was imposed in the respective partitions for the BI and ML inferences. BI analyses were run with 1 million generations using the Monte Carlo Markov chain (MCMC) algorithm. Trees were sampled every 1,000 generations, and 25% of the generations were discarded as burn in once stability in the likelihood values was attained. A half-compatible consensus Bayesian tree was computed from the 750 posterior probability saved trees. ML analysis was computed using 20 starting trees from 20 distinct randomized MP trees and 1,000 bootstrap replicates. MP analysis was based on heuristic searches of 10,000 random-order-entry trees, with tree bisection reconstruction (TBR) branch swapping and saving no more than 10 trees per replicate. The most parsimonious trees were used to compute the respective strict-consensus trees. Branch support was estimated through 1,000 bootstrap replicates (81) using the TBR-M (tree bisection reconstruction swapping, MULPARS off) strategy (82) as a method to reduce computational time. Strains with bootstrap support (BS) values of 75 to 100% or posterior probability support (PPS) values of 90 to 100% were considered moderately to strongly supported.

**(iv) Acyclovir sensitivity testing.** Subjects A and B received valacyclovir oral therapy after HSV infection was diagnosed and after each subsequent episode of disease. Sequence changes were also observed in the polymerase (U<sub>L</sub>30 gene) of subject B. We therefore determined whether any alteration in the isolates' sensitivity to the drug had occurred. All four samples were used to infect Vero cell monolayers in 24-well plates at an MOI of 0.01. After 1 h of adsorption, the wells were washed with PBS, and replicate wells were incubated in the presence of Dulbecco's modified Eagle's medium plus 2% newborn calf serum or supplemented to contain ACV at a concentration of 1.08, 3.6, 12, or 40  $\mu$ M. The well contents were collected 44 h postinfection by scraping, and virus titers were determined by standard plaque assay.

**Accession number(s).** Sequence data (BAM files and FASTA files) were submitted to GenBank under the accession numbers listed in Table 3.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00942-17>.

**SUPPLEMENTAL FILE 1**, PDF file, 0.7 MB.

## ACKNOWLEDGMENTS

We thank Paul Cliften for helpful advice and discussion and Juan A. Villa for helpful discussions about Phyre2 and the PyMOL Molecular Graphics System.

We gratefully acknowledge funding from the NIH Division of Microbiology and Infectious Diseases contracts HHSN272200800003C to R.B.B. and HHSN272201300021I to R.B.B. and L.A.M. and from the Pershing Trust and institutional funds to L.A.M.

## REFERENCES

- Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE, Roizman B, Studdert MJ, Thiry E. 2009. The order Herpesvirales. *Arch Virol* 154:171–177. <https://doi.org/10.1007/s00705-008-0278-4>.
- Roizman B, Sears E. 1996. Herpes simplex viruses and their replication, p 1043–1107. *In* Fields BN, Knipe DM, Howley PM (ed), *Fundamental virology*, 3rd ed. Lippincott-Raven, Philadelphia, PA.
- Taylor TJ, Brockman MA, McNamee EE, Knipe DM. 2002. Herpes simplex virus. *Front Biosci* 7:d752–d764.
- Whitley RJ, Roizman B. 2001. Herpes simplex virus infections. *Lancet* 357:1513–1518. [https://doi.org/10.1016/S0140-6736\(00\)04638-9](https://doi.org/10.1016/S0140-6736(00)04638-9).
- Looker KJ, Magaret AS, Turner KM, Vickerman P, Gottlieb SL, Newman LM. 2015. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PLoS One* 10:e114989. <https://doi.org/10.1371/journal.pone.0114989>.
- Xu F, Schillinger JA, Sternberg MR, Johnson RE, Lee FK, Nahmias AJ, Markowitz LE. 2002. Seroprevalence and coinfection with herpes simplex virus type 1 and type 2 in the United States, 1988–1994. *J Infect Dis* 185:1019–1024. <https://doi.org/10.1086/340041>.
- Bradley H, Markowitz LE, Gibson T, McQuillan GM. 2014. Seroprevalence of herpes simplex virus types 1 and 2—United States, 1999–2010. *J Infect Dis* 209:325–333. <https://doi.org/10.1093/infdis/jit458>.
- Gupta R, Warren T, Wald A. 2007. Genital herpes. *Lancet* 370:2127–2137. [https://doi.org/10.1016/S0140-6736\(07\)61908-4](https://doi.org/10.1016/S0140-6736(07)61908-4).
- Whitley R, Kimberlin DW, Prober CG. 2007. Pathogenesis and disease. Chapter 32. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), *Human herpesviruses: biology, therapy, and immunoprophylaxis*. Cambridge University Press, Cambridge, United Kingdom.
- Wald A, Link K. 2002. Risk of human immunodeficiency virus infection in herpes simplex virus type 2-seropositive persons: a meta-analysis. *J Infect Dis* 185:45–52. <https://doi.org/10.1086/338231>.
- Freeman EE, Weiss HA, Glynn JR, Cross PL, Whitworth JA, Hayes RJ. 2006. Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies. *AIDS* 20:73–83. <https://doi.org/10.1097/01.aids.0000198081.09337.a7>.
- Roizman B, Knipe DM, Whitley RJ. 2013. Herpes simplex viruses, p 1823–1897. *In* Knipe DM, Howley P (ed), *Fields virology*, 6th ed. Lippincott Williams and Wilkins, Philadelphia, PA.
- Roizman B, Knipe DM. 2001. Herpes simplex viruses and their replication, p 2399–2459. *In* Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MM, Roizman B, Straus SE (ed), *Fields virology*, 4th ed. Lippincott, Williams and Wilkins, Philadelphia, PA.
- Braun DK, Batterson W, Roizman B. 1984. Identification and genetic mapping of a herpes simplex virus capsid protein that binds DNA. *J Virol* 50:645–648.
- De Bruyne T, Pieters L, Witvrouw M, De Clercq E, Vanden Berghe D, Vlietinck AJ. 1999. Biological evaluation of proanthocyanidin dimers and related polyphenols. *J Nat Prod* 62:954–958. <https://doi.org/10.1021/np980481o>.
- Gupta R, Hill EL, McClernon D, Davis G, Selke S, Corey L, Wald A. 2005. Acyclovir sensitivity of sequential herpes simplex virus type 2 isolates from the genital mucosa of immunocompetent women. *J Infect Dis* 192:1102–1107. <https://doi.org/10.1086/432766>.
- Tyring SK, Baker D, Snowden W. 2002. Valacyclovir for herpes simplex virus infection: long-term safety and sustained efficacy after 20 years' experience with acyclovir. *J Infect Dis* 186(Suppl 1):S40–S46. <https://doi.org/10.1086/342966>.
- Johnston C, Saracino M, Kuntz S, Magaret A, Selke S, Huang ML, Schiffer JT, Koelle DM, Corey L, Wald A. 2012. Standard-dose and high-dose daily antiviral therapy for short episodes of genital HSV-2 reactivation: three randomised, open-label, cross-over trials. *Lancet* 379:641–647. [https://doi.org/10.1016/S0140-6736\(11\)61750-9](https://doi.org/10.1016/S0140-6736(11)61750-9).
- Field HJ, Vere Hodge RA. 2013. Recent developments in anti-herpesvirus drugs. *Br Med Bull* 106:213–249. <https://doi.org/10.1093/bmb/ldt011>.
- Bacon TH, Levin MJ, Leary JJ, Sarisky RT, Sutton D. 2003. Herpes simplex virus resistance to acyclovir and penciclovir after two decades of antiviral therapy. *Clin Microbiol Rev* 16:114–128. <https://doi.org/10.1128/CMR.16.1.114-128.2003>.
- Gilbert C, Bestman-Smith J, Boivin G. 2002. Resistance of herpesviruses to antiviral drugs: clinical impacts and molecular mechanisms. *Drug Resist Updat* 5:88–114. [https://doi.org/10.1016/S1368-7646\(02\)00021-3](https://doi.org/10.1016/S1368-7646(02)00021-3).
- Andrei G, Snoeck R. 2013. Herpes simplex virus drug-resistance: new mutations and insights. *Curr Opin Infect Dis* 26:551–560. <https://doi.org/10.1097/QCO.0000000000000015>.
- Awasthi S, Shaw C, Friedman H. 2014. Improving immunogenicity and efficacy of vaccines for genital herpes containing herpes simplex virus glycoprotein D. *Expert Rev Vaccines* 13:1475–1488. <https://doi.org/10.1586/14760584.2014.951336>.
- Stanberry LR, Spruance SL, Cunningham AL, Bernstein DI, Mindel A, Sacks S, Tyring S, Aoki FY, Slaoui M, Denis M, Vandepapeliere P, Dubin G. 2002. Glycoprotein-D-adjunct vaccine to prevent genital herpes. *N Engl J Med* 347:1652–1661. <https://doi.org/10.1056/NEJMoa011915>.
- Belshe RB, Leone PA, Bernstein DI, Wald A, Levin MJ, Stapleton JT, Gorfinkel I, Morrow RLA, Ewell MG, Stokes-Riner A, Dubin G, Heineman TC, Schulte JM, Deal CD. 2012. Efficacy results of a trial of a herpes

- simplex vaccine. *N Engl J Med* 366:34–43. <https://doi.org/10.1056/NEJMoa1103151>.
26. Becker Y, Dym H, Sarov I. 1968. Herpes simplex virus DNA. *Virology* 36:184–194. [https://doi.org/10.1016/0042-6822\(68\)90135-9](https://doi.org/10.1016/0042-6822(68)90135-9).
  27. Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ. 1998. The genome sequence of herpes simplex virus type 2. *J Virol* 72:2010–2021.
  28. Kieff ED, Bachenheimer SL, Roizman B. 1971. Size, composition, and structure of the deoxyribonucleic acid of herpes simplex virus subtypes 1 and 2. *J Virol* 8:125–132.
  29. Wadsworth S, Jacob RJ, Roizman B. 1975. Anatomy of herpes simplex virus DNA. II. Size, composition, and arrangement of inverted terminal repetitions. *J Virol* 15:1487–1497.
  30. Newman RM, Lamers SL, Weiner B, Ray SC, Colgrove RC, Diaz F, Jing L, Wang K, Saif S, Young S, Henn M, Laeyendecker O, Tobian AA, Cohen JI, Koelle DM, Quinn TC, Knipe DM. 2015. Genome sequencing and analysis of geographically diverse clinical isolates of herpes simplex virus 2. *J Virol* 89:8219–8232. <https://doi.org/10.1128/JVI.01303-15>.
  31. Kolb AW, Larsen IV, Cuellar JA, Brandt CR. 2015. Genomic, phylogenetic, and recombinational characterization of herpes simplex virus 2 strains. *J Virol* 89:6427–6434. <https://doi.org/10.1128/JVI.00416-15>.
  32. Lai W, Chen CY, Morse SA, Htun Y, Fehler HG, Liu H, Ballard RC. 2003. Increasing relative prevalence of HSV-2 infection among men with genital ulcers from a mining community in South Africa. *Sex Transm Infect* 79:202–207. <https://doi.org/10.1136/sti.79.3.202>.
  33. Colgrove R, Diaz F, Newman R, Saif S, Shea T, Young S, Henn M, Knipe DM. 2014. Genomic sequences of a low passage herpes simplex virus 2 clinical isolate and its plaque-purified derivative strain. *Virology* 450–451:140–145. <https://doi.org/10.1016/j.virol.2013.12.014>.
  34. Dudek TE, Torres-Lopez E, Crumpacker C, Knipe DM. 2011. Evidence for differences in immunologic and pathogenesis properties of herpes simplex virus 2 strains from the United States and South Africa. *J Infect Dis* 203:1434–1441. <https://doi.org/10.1093/infdis/jir047>.
  35. Everett RD, Fenwick ML. 1990. Comparative DNA sequence analysis of the host shutoff genes of different strains of herpes simplex virus: type 2 strain HG52 encodes a truncated UL41 product. *J Gen Virol* 71:1387–1390. <https://doi.org/10.1099/0022-1317-71-6-1387>.
  36. Subak-Sharpe JH, Al-Saadi SA, Clements GB. 1984. Herpes simplex virus type 2 establishes latency in the mouse footpad and in the sensory ganglia. *J Invest Dermatol* 83:67s–71s.
  37. Taha MY, Brown SM, Clements GB. 1988. Neurovirulence of individual plaque stocks of herpes simplex virus type 2 strain HG 52. *Arch Virol* 103:15–25. <https://doi.org/10.1007/BF01319805>.
  38. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, Enquist LW, Legendre M, Davison AJ. 2014. Evolution and diversity in human herpes simplex virus genomes. *J Virol* 88:1209–1227. <https://doi.org/10.1128/JVI.01987-13>.
  39. Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626. <https://doi.org/10.1038/217624a0>.
  40. King JL, Jukes TH. 1969. Non-Darwinian evolution. *Science* 164:788–798. <https://doi.org/10.1126/science.164.3881.788>.
  41. Lodowski DT, Barnhill JF, Pyskadlo RM, Ghirlando R, Sterne-Marr R, Tesmer JJG. 2005. The role of G beta gamma and domain interfaces in the activation of G protein-coupled receptor kinase 2. *Biochemistry* 44:6958–6970. <https://doi.org/10.1021/bi050119q>.
  42. Metrick CM, Chadha P, Heldwein EE. 2015. The unusual fold of herpes simplex virus 1 UL21, a multifunctional tegument protein. *J Virol* 89:2979–2984. <https://doi.org/10.1128/JVI.03516-14>.
  43. Takakuwa H, Goshima F, Koshizuka T, Murata T, Daikoku T, Nishiyama Y. 2001. Herpes simplex virus encodes a virion-associated protein which promotes long cellular processes in over-expressing cells. *Genes Cells* 6:955–966. <https://doi.org/10.1046/j.1365-2443.2001.00475.x>.
  44. Metrick CM, Heldwein EE. 2016. Novel structure and unexpected RNA-binding ability of the C-terminal domain of herpes simplex virus 1 tegument protein UL21. *J Virol* 90:5759–5769. <https://doi.org/10.1128/JVI.00475-16>.
  45. Liu S, Knafels JD, Chang JS, Waszak GA, Baldwin ET, Deibel MR, Jr, Thomsen DR, Homa FL, Wells PA, Tory MC, Poorman RA, Gao H, Qiu X, Seddon AP. 2006. Crystal structure of the herpes simplex virus 1 DNA polymerase. *J Biol Chem* 281:18193–18200. <https://doi.org/10.1074/jbc.M602414200>.
  46. Pitts JD, Klabis J, Richards AL, Smith GA, Heldwein EE. 2014. Crystal structure of the herpesvirus inner tegument protein UL37 supports its essential role in control of viral trafficking. *J Virol* 88:5462–5473. <https://doi.org/10.1128/JVI.00163-14>.
  47. Echeverria N, Moratorio G, Cristina J, Moreno P. 2015. Hepatitis C virus genetic variability and evolution. *World J Hepatol* 7:831–845. <https://doi.org/10.4254/wjh.v7.i6.831>.
  48. Kolb AW, Adams M, Cabot EL, Craven M, Brandt CR. 2011. Multiplex sequencing of seven ocular herpes simplex virus type-1 genomes: phylogeny, sequence variability, and SNP distribution. *Invest Ophthalmol Vis Sci* 52:9061–9073. <https://doi.org/10.1167/iov.11-7812>.
  49. Norberg P, Bergstrom T, Rekabdar E, Lindh M, Liljeqvist JA. 2004. Phylogenetic analysis of clinical herpes simplex virus type 1 isolates identified three genetic groups and recombinant viruses. *J Virol* 78:10755–10764. <https://doi.org/10.1128/JVI.78.19.10755-10764.2004>.
  50. Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL. 2014. Evolutionary origins of human herpes simplex viruses 1 and 2. *Mol Biol Evol* 31:2356–2364. <https://doi.org/10.1093/molbev/msu185>.
  51. Capra JA, Singh M. 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics* 23:1875–1882. <https://doi.org/10.1093/bioinformatics/btm270>.
  52. Yates CM, Filippis I, Kelley LA, Sternberg MJ. 2014. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol* 426:2692–2701. <https://doi.org/10.1016/j.jmb.2014.04.026>.
  53. Cano-Monreal GL, Wylie KM, Cao F, Tavis JE, Morrison LA. 2009. Herpes simplex virus 2 UL13 protein kinase disrupts nuclear lamins. *Virology* 392:137–147. <https://doi.org/10.1016/j.virol.2009.06.051>.
  54. Overton HA, McMillan DJ, Klavinskis LS, Hope L, Ritchie AJ, Wong-Kai-In P. 1992. Herpes simplex virus type 1 gene UL13 encodes a phosphoprotein that is a component of the virion. *Virology* 190:184–192. [https://doi.org/10.1016/0042-6822\(92\)91204-8](https://doi.org/10.1016/0042-6822(92)91204-8).
  55. Kawaguchi Y, Kato K. 2003. Protein kinases conserved in herpesviruses potentially share a function mimicking the cellular protein kinase cdc2. *Rev Med Virol* 13:331–340. <https://doi.org/10.1002/rmv.402>.
  56. Morrison EE, Wang YF, Meredith DM. 1998. Phosphorylation of structural components promotes dissociation of the herpes simplex virus type 1 tegument. *J Virol* 72:7108–7114.
  57. Purves FC, Roizman B. 1992. The UL13 gene of herpes simplex virus 1 encodes the functions for posttranslational processing associated with phosphorylation of the regulatory protein alpha 22. *Proc Natl Acad Sci U S A* 89:7310–7314. <https://doi.org/10.1073/pnas.89.16.7310>.
  58. Tanaka M, Nishiyama Y, Sata T, Kawaguchi Y. 2005. The role of protein kinase activity expressed by the UL13 gene of herpes simplex virus 1: the activity is not essential for optimal expression of UL41 and ICP0. *Virology* 341:301–312. <https://doi.org/10.1016/j.virol.2005.07.010>.
  59. Advani SJ, Brandimarti R, Weichselbaum RR, Roizman B. 2000. The disappearance of cyclins A and B and the increase in activity of the G(2)/M-phase cellular kinase cdc2 in herpes simplex virus 1-infected cells require expression of the alpha22/U(S)1.5 and U(L)13 viral genes. *J Virol* 74:8–15. <https://doi.org/10.1128/JVI.74.1.8-15.2000>.
  60. Smith RF, Smith TF. 1989. Identification of new protein kinase-related genes in three herpesviruses, herpes simplex virus, varicella-zoster virus, and Epstein-Barr virus. *J Virol* 63:450–455.
  61. Coulter LJ, Moss HW, Lang J, McGeoch DJ. 1993. A mutant of herpes simplex virus type 1 in which the UL13 protein kinase gene is disrupted. *J Gen Virol* 74:387–395. <https://doi.org/10.1099/0022-1317-74-3-387>.
  62. Long MC, Leong V, Schaffer PA, Spencer CA, Rice SA. 1999. ICP22 and the UL13 protein kinase are both required for herpes simplex virus-induced modification of the large subunit of RNA polymerase II. *J Virol* 73:5593–5604.
  63. Desai P, Sexton GL, Huang E, Person S. 2008. Localization of herpes simplex virus type 1 UL37 in the Golgi complex requires UL36 but not capsid structures. *J Virol* 82:11354–11361. <https://doi.org/10.1128/JVI.00956-08>.
  64. Bucks MA, Murphy MA, O'Regan KJ, Courtney RJ. 2011. Identification of interaction domains within the UL37 tegument protein of herpes simplex virus type 1. *Virology* 416:42–53. <https://doi.org/10.1016/j.virol.2011.04.018>.
  65. Kelly BJ, Diefenbach E, Fraefel C, Diefenbach RJ. 2012. Identification of host cell proteins which interact with herpes simplex virus type 1 tegument protein pUL37. *Biochem Biophys Res Commun* 417:961–965. <https://doi.org/10.1016/j.bbrc.2011.12.044>.
  66. Ko DH, Cunningham AL, Diefenbach RJ. 2010. The major determinant for addition of tegument protein pUL48 (VP16) to capsids in herpes simplex

- virus type 1 is the presence of the major tegument protein pUL36 (VP1/2). *J Virol* 84:1397–1405. <https://doi.org/10.1128/JVI.01721-09>.
67. Jambunathan N, Chouljenko D, Desai P, Charles AS, Subramanian R, Chouljenko VN, Kousoulas KG. 2014. Herpes simplex virus 1 protein UL37 interacts with viral glycoprotein gK and membrane protein UL20 and functions in cytoplasmic virion envelopment. *J Virol* 88:5927–5935. <https://doi.org/10.1128/JVI.00278-14>.
  68. Chouljenko DV, Jambunathan N, Chouljenko VN, Naderi M, Brylinski M, Caskey JR, Kousoulas KG. 2016. Herpes simplex virus 1 UL37 protein tyrosine residues conserved among all alphaherpesviruses are required for interactions with glycoprotein K, cytoplasmic virion envelopment, and infectious virus production. *J Virol* 90:10351–10361. <https://doi.org/10.1128/JVI.01202-16>.
  69. 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.
  70. Thorvaldsdottir 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>.
  71. 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>.
  72. Ray A, Lindahl E, Wallner B. 2012. Improved model quality assessment using ProQ2. *BMC Bioinformatics* 13:224. <https://doi.org/10.1186/1471-2105-13-224>.
  73. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407–415. <https://doi.org/10.1093/bioinformatics/btg427>.
  74. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>.
  75. Swofford DL. 2003. PAUP\*: phylogenetic analysis using parsimony (and other methods) 4.0.b10. Sinauer Associates, Sunderland, MA.
  76. Akaike H. 1998. A new look at the statistical model identification, p 215–222. *In* Parzen E, Tanabe K, Kitagawa G (ed), *Selected papers of Hirotugu Akaike*. Springer, New York, NY.
  77. Schwarz G. 1978. Estimating the dimension of a model, p 461–464. Institute of Mathematical Statistics, Bethesda MD.
  78. Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst Biol* 52:674–683. <https://doi.org/10.1080/10635150390235494>.
  79. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704. <https://doi.org/10.1080/10635150390235520>.
  80. Darrriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <https://doi.org/10.1038/nmeth.2109>.
  81. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.
  82. DeBry RW, Olmstead RG. 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Syst Biol* 49:171–179. <https://doi.org/10.1080/10635150050207465>.
  83. 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. <https://doi.org/10.1186/gb-2009-10-3-r25>.
  84. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
  85. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25:1968–1969. <https://doi.org/10.1093/bioinformatics/btp347>.
  86. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
  87. Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28:2184–2185. <https://doi.org/10.1093/bioinformatics/bts356>.
  88. 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 (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.