Virology 450-451 (2014) 140–145

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Genomic sequences of a low passage herpes simplex virus 2 clinical isolate and its plaque-purified derivative strain

Robert Colgrove [a,1], Fernando Diaz [a], Ruchi Newman [b], Sakina Saif [b], Terry Shea [b], Sarah Young [b], Matt Henn [b,2], David M. Knipe [a,*]

[a] Department of Microbiology and Immunobiology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, United States
[b] Broad Institute, Cambridge, MA, United States

ARTICLE INFO

Article history:
Received 24 October 2013
Returned to author for revisions
12 November 2013
Accepted 9 December 2013
Available online 31 December 2013

Keywords:
Genital herpes
Herpes simplex virus 2
DNA virus genome
Sexually transmitted disease

ABSTRACT

Herpes simplex virus 2 is an important human pathogen as the causative agent of genital herpes, neonatal herpes, and increased risk of HIV acquisition and transmission. Nevertheless, the only genomic sequence that has been completed is the attenuated HSV-2 HG52 laboratory strain. In this study we defined the genomic sequence of the HSV-2 SD90e low passage clinical isolate and a plaque-purified derivative, SD90-3P. We found minimal sequence differences between SD90e and SD90-3P. However, in comparisons with the HSV-2 HG52 reference genome sequence, the SD90e genome ORFs contained numerous point mutations, 13 insertions/deletions (indels), and 9 short compensatory frameshifts. The indels were true sequence differences, but the compensatory frameshifts were likely sequence errors in the original HG52 sequence. Because HG52 virus is less virulent than other HSV-2 strains and may not be representative of wildtype HSV-2 strains, we propose that the HSV-2 SD90e genome serve as the new HSV-2 reference genome.

Introduction

Herpes simplex virus 2 (HSV-2), the major causative agent of genital herpes, undergoes a primary infection in the genital mucosa and spreads to sacral ganglia where the virus establishes a latent infection in sensory neurons. Reactivation of the latent virus leads to recurrent genital infections and lesions, called genital herpes. HSV-2 causes life-threatening infections in neonates infected during delivery and increases the risk of HIV acquisition and transmission (Roizman et al., 2013). Therefore, HSV-2 is an important human pathogen, and additional antivirals and a vaccine are needed.

The herpes simplex viruses are large double-stranded DNA (dsDNA) viruses that replicate in the nuclei of host cells (Roizman et al., 2013). The HSV genome is a linear dsDNA molecule comprised of two covalently linked segments, the long (L) and short (S) segments, which each consist of unique sequences (U_L and U_S) bounded by inverted repeats (Hayward et al., 1975) (Fig. 1). Several HSV-1 genomic sequences have been determined (Macdonald et al., 2012a, b; McGeoch et al., 1988; McGeoch et al., 1986; Szpara et al., 2010), but the complete genome of only one strain of HSV-2 has been determined. The only available HSV-2 genomic sequence, that of the HG52 laboratory strain (Timbury, 1971), was determined a number of years ago (Dolan et al., 1998) and defined the genome size as 154,476 basepairs (bps). The HG52 DNA sequence was recently updated by Andrew Davison and provided in GenBank (JN561323.1), but the original sequence still serves as the reference genome for HSV-2. Although there is no complete genome sequence of a low-passage clinical isolate of HSV-2, sequencing of specific genes of HSV isolates gives evidence of at least limited sequence diversity among HSV-2 isolates (Norberg et al., 2007). Sequence analysis of glycoprotein genes in African and European HSV-2 isolates showed limited sequence diversity and defined two possible genogroups, one consisting entirely of African isolates and another containing both European and African isolates (Norberg et al., 2007). In addition, laboratory strains of HSV-2 show polymorphisms in restriction endonuclease cleavage sites (Hayward et al., 1975), and there is evidence for differences in immunological and pathogenic properties of HSV-2 strains from the United States and South Africa (Dudek et al., 2011).

The NCBI reference sequence strain, HSV-2 strain HG52 (Accession number NC_001798), shows variable virulence, in that plaque stocks showed LD_{50} values ranging from greater than 10^5 PFU to less than 10^3 PFU by cranial inoculation (Taha et al., 1988). When inoculated peripherally, HSV-2 HG52 seems to be relatively avirulent. Very limited paralysis was observed when HG52 was inoculated in mice by the footpad route (Subak-Sharpe et al., 1984). In a study involving corneal infection of mice, 10^6–10^7 PFU of HG52 showed no

* Corresponding author.
E-mail address: david_knipe@hms.harvard.edu (D.M. Knipe).
[1] Current address: Mount Auburn Hospital, Cambridge, MA, United States.
[2] Current address: Seres Health, Cambridge, MA, United States.

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.12.014

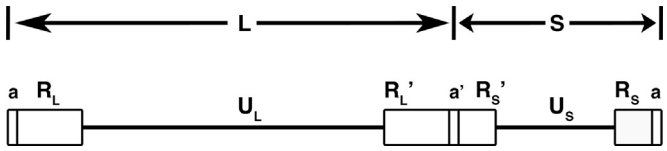

Fig. 1. Diagram of the structure of the Herpes simplex virus genome. The top row shows the long (L) component and the short (S) component of the HSV genome. The bottom row shows the unique sequences as a line and the boxes denote the repeated sequences. U_L=unique long component sequences; U_S=unique short component sequences; R_L and R_L'=inverted repeats bounding the long component; R_S and R_S' denote inverted repeats bounding the S component. a=terminal repeat also located at the L/S junction.

mortality while similar dosages of HSV-2 strains 333 or 186 showed 100% mortality (Mitchell et al., 1990). The LD_{50} for other HSV-2 strains is in the range of $1–6 \times 10^3$ PFU via the vaginal route (Blakeney et al., 2005; Dudek et al., 2011). Furthermore, unlike other HSV-2 strains, HSV-2 HG52 is unable to shut off host protein synthesis. This defect is likely due to a frameshift mutation in the *U_L41* gene, which encodes the viral host shutoff protein (Everett and Fenwick, 1990). Despite the knowledge of the host shutoff defect, the full genotype of HG52 responsible for the attenuated phenotype is currently not known.

In total, these considerations highlight the urgent need for a complete genome sequence of a low passage isolate of HSV-2. In this study we have generated the complete genome sequence of an early passage virulent African HSV-2 clinical isolate, SD90e, as well as a plaque-purified version of this isolate. We propose that this new clinical isolate genome could serve as a new HSV-2 reference genome.

Results

Comparison of an HSV-2 early passage genome with the HG52 reference genome.

The HSV-2 SD90 virus, which was isolated at an STD clinic in South Africa (Lai et al., 2003), has been passaged only a limited number of times in cell culture and shows high virulence in a mouse genital model (Dudek et al., 2011), similar to most other clinical HSV-2 strains (Blakeney et al., 2005; Dudek et al., 2011). We determined the genomic sequence for an SD90 early passage isolate, named SD90e, using viral DNA purified from infected cell lysates by double banding on sodium iodide (NaI) density gradients. Illumina sequencing of this sample generated ~6000-fold sequence coverage. Assembly of Illumina data using de novo assembly programs gave a number of blocks of sequence, or contigs, as observed for HSV-1 (Szpara et al., 2010). To refine the assembly and fill in large gaps, we mapped the sequence reads to the individual U_L, U_S, R_L and R_S regions containing 200 bp of overlapping sequence based on the HG52 reference sequence. This gave 5 large and 3 small contigs. The resulting contigs were used to manually reconstruct a full genome sequence by (1) mining Illumina raw reads when necessary to connect the contigs, (2) generating R_L and R_S inverted repeats bounding their respective unique sequences by inverting a copy of each sequence, and (3) assignment of numbers of repetitive sequences (*a* sequences) at the termini and the junction of R_L and R_S to match the HG52 reference sequence. There was one region in each of the *ICP4*/*R_S1*, *ICP34.5*/*R_L1*, *U_L36*, and *ICP0*/R_L2 *ORFs* whose sequence was ambiguous. We therefore used the primers listed in Table 1 and the conditions described in “Materials and methods” to PCR-amplify these regions from SD90e DNA, and we then sequenced the amplified DNA by conventional Sanger sequencing. The sequences in the repeated regions in the *ICP4*/*R_S1*, *ICP34.5*/*R_L1*, and *ICP0*/R_L2 *ORFs* were determined using the primers in Table 1.

The *U_L36* gene repeats were more refractory in that Sanger sequencing of the SD90e fragment showed at least 9 copies of the 15 nucleotide repeat (caggggcggctgggg) before the sequence appeared to move into unique sequences, but the sequence was not unequivocal. Parallel sequencing of our isolate of HSV-2 HG52 DNA showed 6 copies of the *U_L36* repeat, in contrast to the 11 copies in the GenBank reference sequence. To attempt to confirm these numbers of repeats, we analyzed the PCR products in an 8% polyacrylamide gel (Fig. 2). The SD90e amplimer was approximately 550 bp, consistent with the 547 bp predicted by 9 copies of the repeated sequence while the HG52 amplimer was approximately 500 bp, consistent with the 502 bp predicted by 6 copies of the repeated sequence. Therefore, we have provisionally described the *U_L36* gene of SD90e as having 9 copies of this repeated sequence, but further analysis is needed to confirm this.

The combination of the Illumina and Sanger sequence analyses gave a complete SD90e genome of 154,096 basepairs (GenBank accession number KF781518). We then compared our SD90e genome assembly with the HSV-2 HG52 reference genome in GenBank (NC_001798) and the recent HG52 sequence submitted by Andrew Davison (GenBank accession number JN561323.1, determined by “sequence of viral DNA and polyA RNA harvested at 10 h after infection”) using the Genome Annotation Transfer Utility (GATU) program (Pickett et al., 2012) from the Virus Pathogen Resource (http://www.viprbrc.org). The SD90e genome contained more than 100 nucleotide differences within open reading frames compared to the HG52 reference sequence, but overall, U_S was 99.3% conserved and U_L was 99.6% conserved between the two genomes.

Nine ORFs contained a combined total of thirteen indels (insertions or deletions) (Table 2). These indels were multiples of 3 nucleotides so the reading frame was not altered. The original HG52 sequence and the more recent Davison HG52 sequence were identical in these regions (Table 2). Furthermore, our own sequence of HG52 shows the same indels (results not shown). Therefore, these indels are likely to be real sequence differences between SD90e and HG52. Because several of the affected ORFs encode essential or important viral proteins, we predict that these changes could account for the reduced virulence of the HG52 virus. For example, the SD90e *ICP4* (*R_S1*) gene encoding the major viral transactivator protein showed three indels, a 3-codon deletion, a 1-codon insertion, and a 6-codon insertion (Table 2). The *R_L2* gene encoding the ICP0 immediate early protein showed a 20-codon deletion. The *U_L27* ORF encoding glycoprotein B, which is essential for viral entry, showed a 3-codon deletion near the amino terminus. The *U_L32* ORF encoding an essential DNA packaging protein showed a 2-codon deletion. The *U_L49* ORF encoding the VP22 virion tegument protein showed a 2-codon insertion. The other ORFs affected encode viral proteins that play important roles *in vivo*. The *R_L1* gene encoding the ICP34.5 antagonist of the cellular interferon response contained an 8-codon deletion and a 1-codon insertion. The *U_L39* ORF encoding the large subunit of ribonucleotide reductase contained a 3-codon deletion and a 2-codon insertion. The *U_S8* ORF, encoding the glycoprotein E subunit of the viral F_c receptor, contained a 3-codon insertion. Finally, the *U_S4* ORF encoding glycoprotein G contained a 1-codon deletion within the HSV-2 specific region. Variation within this region could affect the ability of the standard HSV-2 reference sera to detect the HSV-2 specific antigen in the C-terminus of gG-2.

In addition, there were 9 sequence differences between the SD90e and HG52 reference sequences that consisted of pairs of closely spaced single nucleotide indels that change the reading frame for a short distance before it goes back into the reference