Sequence-Based Methods for Identifying Epidemiologically Linked Herpes Simplex Virus Type 2 Strains

Emily Toth Martin, David M. Koelle, Benjamin Byrd, Meei-Li Huang, Jeffrey Vieira, and Anna Wald

Department of Epidemiology and Department of Pathobiology, School of Public Health and Community Medicine, and Department of Medicine and Department of Laboratory Medicine, School of Medicine, University of Washington, and Program in Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, Washington

Received 10 January 2006/Returned for modification 15 March 2006/Accepted 1 May 2006

Traditional methods for confirming the identity of herpes simplex virus (HSV) isolates use restriction fragment length polymorphism (RFLP). However, RFLP is less amenable to high-throughput analyses of many samples, and the extent to which small differences in RFLP patterns distinguish between different viral strains remains unclear. Viral HSV type 2 (HSV-2) DNA isolates from 14 persons experiencing a primary HSV-2 infection and from their sexual partners were analyzed by RFLP and heteroduplex mobility assays. We also compared the HSV-2 sequences from seven regions, including noncoding regions between UL19 and UL20, UL24 and UL25, UL37 and UL38, and UL41 and UL42 and coding segments of the gC, gB, and gG genes. Although the resulting RFLP patterns of the couples were almost identical, minor banding differences existed between the source and susceptible partners in five couples. Heteroduplex mobility assays were unable to distinguish between unrelated strains. Overall, 22 sites of sequence variation were found in 1,482 bp of analyzed sequence. The DNA sequences differentiated between all unrelated infections, and epidemiologically related isolates had identical sequences in all but two pairs. Our results suggest that a multifocus assay based on several DNA sequences has the potential to be an informative tool for identifying epidemiologically related HSV-2 strains.

Epidemiologic investigations of human herpes simplex virus type 2 (HSV-2) have relied on the use of restriction fragment length polymorphisms (RFLP) to demonstrate superinfection and distinguish between unrelated strains (5, 15, 19). In the past several years, molecular techniques, including heteroduplex mobility assays (HMA) and sequencing, have superseded RFLP for distinguishing other viruses such as hepatitis C virus and human immunodeficiency virus (1, 3, 22, 23). Recently, Norberg et al. (16) showed the utility of HSV-1 DNA sequencing for molecular epidemiology at a population and an individual level through sequencing the near-adjacent coding regions of HSV-1 glycoprotein G (gG), gE, and gI. For HSV-1, sequencing of hypervariable regions ReIV and ReVII has been shown to be useful for discriminating between isolates (14). Targeted DNA sequencing of HSV-1 hypervariable regions can also be used to detect a second exogenous HSV-1 infection (17). However, the presence of hypervariable regions in the HSV-2 genome has not been definitively established.

To develop sequence-based methods for distinguishing between HSV-2 strains, we performed RFLP, HMA, and DNA sequence analysis of selected regions of the HSV-2 genome. We attempted to maximize the degree of sequence variability through our selection of genome regions for study. Noncoding genome regions were chosen for sequencing because we hypothesized a lack of purifying selection in these regions, and glycoprotein-coding regions were chosen based on previous sequence variation data in analogous segments in the HSV-1 genome (4, 13, 16). Our goals were to distinguish between strains from unrelated individuals and to identify epidemiologically linked HSV-2 strains.

(Materials and Methods)

Sample collection. Swabs from genital herpes lesions were collected as part of an ongoing study of HSV-2 transmission conducted at the University of Washington Virology Research Clinic in Seattle. Informed consent was obtained from patients, and human experimentation guidelines of the University of Washington were followed in the conduct of research with human subjects. Transmission events were identified by clinical presentation and documented by viral culture and seroconversion to HSV-2 by Western blotting (2). The HSV-2 samples were isolated on HDF (human tonsil) cells and then passed once in Vero cells, and DNA from the cytoplasm fraction was isolated using phenol-chloroform extraction as previously described (8). Virus was purified from the cytoplasm fraction to reduce the amount of human genomic DNA background. Whole-genome amplification was used to increase stocks of purified genomic HSV-2 DNA. This procedure was performed by multiple displacement amplification (MDA) (6) using a Repli-g kit including high-fidelity DNA polymerase and random exonuclease-resistant primers (Molecular Staging, New Haven, CT).

RFLP. Approximately 750 ng of HSV-2 DNA from each individual was digested to completion with Sall and separately with Xhol (Invitrogen, Carlsbad, CA) by methods similar to those reported previously (12). Digests were then electrophoresed on a 0.6% agarose gel overnight. DNA fragments were detected with SYBR green.

PCR. Seven segments of the HSV-2 genome (Table 1) were amplified by PCR from each isolate. Segments gC, noncoding region 1 (NC1), and NC3 were amplified using the Stratagene (La Jolla, CA) PfuUltra enzyme. Segments from gB, gG, NC2, and NC4 were amplified using the Applied Biosystems (Foster City, CA) AmpliTaq enzyme when they failed to amplify efficiently with the PfuUltra enzyme. PCRs included 8% glycerol to increase reaction efficiency, 0.3

* Corresponding author. Mailing address: Virology Research Clinic, University of Washington, 600 Broadway, Suite 400, Seattle, WA 98122. Phone: (206) 720-4340. Fax: (206) 720-4371. E-mail: annawald@u.washington.edu.
μM of each primer (Table 1), and 10^6 to 10^7 copies of DNA template based on standard quantitative real-time TaqMan PCR (Applied Biosystems, Foster City, CA) (18). Because of high GC content in the regions amplified, 5% dimethyl sulfoxide was added to the PCRs for all regions except NC1. SYBR green in dimethyl sulfoxide and ROX (Invitrogen, Carlsbad, CA) were added to enable real-time detection of amplicons. Three-step PCR cycling conditions included denaturing at 94°C for 20 s, annealing at 68°C for 1 min for NC1 and NC3 and at 60°C for 30 s for GG, gB, gC, NC2, and NC4, and extension at 72°C for 1 min for 35 cycles. Amplicons were then purified using DNA Clean & Concentrator (Zymo Research, Orange, CA).

DNA sequencing. All PCR amplicons were routinely sequenced bidirectionally, except for that in region NC3, using BigDye Terminator (Applied Biosystems, Foster City, CA). Second-strand sequencing of NC3 was done only if polymorphisms were suspected on unidirectional sequencing. Sequencing primers were identical to those used for PCR (Table 1). The sequences were analyzed using Sequencher, version 4.1.4 (Gene Codes, Ann Arbor, MI). HSV-2 strain HG52 (GenBank accession no. NC001798) was used as the reference sequence. The sequences were aligned, and polymorphisms were detected using Sequencher, version 4.1.4 (Gene Codes, Ann Arbor, MI). HSV-2 strain HG52 (GenBank accession no. NC001798) was used as the reference sequence. The sequences were analyzed using Sequencher, version 4.1.4 (Gene Codes, Ann Arbor, MI).

TABLE 1. Regions of HSV-2 genome amplified by PCR

<table>
<thead>
<tr>
<th>Region</th>
<th>ORFa</th>
<th>HSV-2 genome locationb</th>
<th>Size (bp)</th>
<th>Left primer</th>
<th>Right primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC1</td>
<td>Between UL19 and UL20</td>
<td>40560–40908</td>
<td>348</td>
<td>5− GCTGAGGGCGGCCGGCGTTC</td>
<td>5− CAGAGGCGCCCTTGGGTTTC</td>
</tr>
<tr>
<td>NC2</td>
<td>Between UL41 and UL42</td>
<td>93273–93464</td>
<td>192</td>
<td>5− CATTCCGGCCCGATGTA</td>
<td>5− CACCAATCCGAGCTGATAG</td>
</tr>
<tr>
<td>NC3</td>
<td>Between UL24 and UL25</td>
<td>48706–49047</td>
<td>342</td>
<td>5− CCGGTGGGACGCGAGACAGA</td>
<td>5− CTGCCAGCGCGCTGAAAGGTTAG</td>
</tr>
<tr>
<td>NC4</td>
<td>Between UL37 and UL38</td>
<td>84732–84981</td>
<td>250</td>
<td>5− TATAGCAGGCGTATGTC</td>
<td>5− ACACGCCACGAAACAC</td>
</tr>
<tr>
<td>gB</td>
<td>UL27</td>
<td>54093–54319</td>
<td>227</td>
<td>5− CAGCATGGGTGGATATCCCTG</td>
<td>5− GCTTTCCGTGATCCAAGCAG</td>
</tr>
<tr>
<td>gC</td>
<td>UL44</td>
<td>97570–97797</td>
<td>228</td>
<td>5− CTCATCATCGAAGAGCGT</td>
<td>5− GCCTCGACACAGACAACT</td>
</tr>
<tr>
<td>gG</td>
<td>US4</td>
<td>138209–138786</td>
<td>168</td>
<td>5− TACAGTACACACGGCTACC</td>
<td>5− GCAGGAAGCATTACAG</td>
</tr>
</tbody>
</table>

a ORF, open reading frame.
b HSV-2 strain HG52 (GenBank accession no. NC001798) coordinates.

RESULTS

Fourteen pairs of sexual partners were included in the analysis. All of the pairs were heterosexual couples, and 5 of the 28 individuals belonged to racial minorities. Each pair included a person who was asked to transmit HSV-2 (partner A) and that person's sexual partner, who was diagnosed with a laboratory-documented newly acquired HSV-2 infection (partner B) thought to be acquired from partner A. Acquisition of initial HSV-2 infection was diagnosed by clinical presentation with genital lesions, HSV-2 isolation from the genital tract, and seroconversion to HSV-2 by Western blotting (2); all three criteria were present for all 14 partner B's.

RFLP. Nine couples had identical RFLP patterns after digestion with each enzyme (Table 2; Fig. 1). Minor banding differences showed no more than a 2-band difference by use of either enzyme. However, comparisons of unrelated individuals showed no more than a 2-band difference could be used as a threshold for determining relatedness with one enzyme. For example, between pairs 2, 4, 7, 8, 13, and 14, the mean number of band differences between unrelated individuals was 3.9, with some unrelated individuals differing by only 1 band (Fig. 1). However, each individual was more closely related to his or her partner than to any unrelated individual.

TABLE 2. Number of band differences within pairs, by restriction enzyme

<table>
<thead>
<tr>
<th>Pair ID</th>
<th>Differences with Sall</th>
<th>Differences with Xhol</th>
<th>Time (no. of days) between collection dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Same day</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>Same day</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>2 bands</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>Same day</td>
</tr>
<tr>
<td>6</td>
<td>2 bands</td>
<td>No data</td>
<td>Same day</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>10</td>
<td>2 bands</td>
<td>2 bands</td>
<td>17</td>
</tr>
<tr>
<td>11</td>
<td>2 bands</td>
<td>2 bands</td>
<td>13</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>13</td>
<td>2 bands</td>
<td>1 bands</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

a ID, identification number.
**HMA.** NC1 amplicons were chosen for HMA analysis because the high yield of the PCR provided ample product for multiple analysis methods. Slight mobility shifts were observed for pairs 2 and 8 and for partner B from pair 1 and partner A from pair 3 (Fig. 2). All other samples had indistinguishable migration. Concurrent DNA sequencing showed the variation between the test strains and the HSV-2 strain 333 probe to be lower than the 1% to 2% sequence divergence previously found to be necessary for measurable differences in heteroduplex mobility between samples (3). The slight band differences found in pairs with identical sequences (pairs 1 and 3) may be due to a low level of variation from the probe strain, resulting in ambiguous HMA profiles (3). HMA was not used to analyze the other regions due to the lack of migration seen using the NC1 region.

**DNA sequencing.** Eight regions of HSV-2 sequence were amplified from each isolate except 3A and 9B. These two isolates did not have amplification at region NC4. For pairs 3 and 11, genotypes for the first two polymorphic sites of NC3 could not be determined due to inefficient sequencing across the poly(C) region of the amplicon.

Noncoding regions showed the most variation. Among all 28 samples, one or more strains showed variation at a total of 22.44 nucleotides/1,000, 30.04 nucleotides/1,000, and 33.33 nucleotides/1,000 for NC1, NC3, and NC4, respectively. No variation was found in 190 bp of NC2. In coding regions, one single-nucleotide polymorphism each was found in 152 bp of gG and 228 bp of gC, and no variation was present in 217 bp of gB. Overall, 21 sites of sequence variation were found in 1,482 bp of analyzed sequence (Table 3). Five of those sites consisted of variable-length repeats of one or two nucleotides. Two of these repeat mutation sites were in NC1, two were in NC2, and one was in NC4 (Table 3). The remaining sites were single-nucleotide polymorphisms (Fig. 3). In addition to the variations listed, one individual had the sequence GGAGAGGGGA in place of a variable-length G repeat located near position 84794 in the NC4 region.

The HG52 sequence differed from the consensus sequence of our samples at two sites in NC1 and one site in NC4 (Fig. 3). In regions with no variation (NC2 and gB), the HG52 sequence was consistent with all other samples. Eleven of the 14 pairs had sequences that were identical between the related partners at all sites. Pair 8 differed at site 48958, a variable-length C repeat in a noncoding region (Table 3). Pair 12 differed at site 84796, a variable-length GT repeat in a noncoding region (Table 3).
length GT repeat in a noncoding region. Pair 10 differed at six sites. A clinical chart review revealed multiple concurrent sexual partners for partner B from pair 10. This sexual history, in combination with the sequence data, indicated that partner A is an unlikely source of partner B’s genital herpes infection.

To assess the stability of the sequences during tissue culture, we evaluated DNA samples from HSV-2 lesions from four individuals before and after culture. All genomic regions were evaluated for specimens from all but one individual. We found no differences in sequence between DNA amplified directly from lesions and DNA obtained after virus was cultured from the same clinical lesions and passaged twice. During this sequencing of additional strains, occasional 3-nucleotide polymorphisms were uncovered, two in the noncoding region NC1.
and one in the noncoding region NC4. These can be accessed from the GenBank submissions accompanying this report. These data are consistent with stability in HSV-2 genomic DNA during a limited number of in vitro passages in these loci and the identification of bona fide polymorphisms in our study.

**DISCUSSION**

We found that sequencing short segments of both coding and noncoding regions of HSV-2 allowed us to correctly identify epidemiologically linked strains of HSV-2 as well as to differentiate between all of the unrelated strains. These results were confirmed by RFLP, which has historically been used to demonstrate strain variation (5, 15, 19). As a result of the low rate of mutations overall, HMA was found to be of limited utility for differentiating HSV-2 strains. A few differences between related strains were observed by both RFLP and DNA sequencing. Among the 13 pairs that are assumed to be epidemiologically related, 3 pairs were found to have DNA sequence differences and 5 pairs were found to have RFLP differences. These results indicate that a minor level of variation can be expected even between individuals related by a transmission event, and therefore, care needs to be taken in the interpretation of minor differences between individuals. Similar minor variations have been reported in strains recovered from consecutive recurrences from individuals over time (5, 12, 15, 21). However, the RFLP method requires visual interpretation to determine the relatedness of samples. DNA sequencing is more suitable for efficient analyses of large numbers of samples than RFLP and may be more amenable to standardization. Therefore, we believe that it is better suited for analysis involving large numbers of HSV-2 strains to assess epidemiological relationships. DNA polymorphisms can be objectively identified and labeled. Specifying the most appropriate set of polymorphisms to use for comparison of HSV-2 isolates can lead to a more rigorous standard for determining relatedness that is more easily communicated than RFLP findings.

Past HSV sequencing efforts have focused on coding regions, particularly glycoprotein sequences (4, 13, 16). Two publications have reported that large sections of HSV-1 sequence, including complete coding regions, could be used for examining strain differences between populations (4) and individuals (16). However, a small segment (179 bp) of the HSV-2 DNA polymerase gene was found to have no variability between two samples and a standard strain (13). Nevertheless, we found that similarly small segments could be of use if they were generated from intergenic spacers that do not code for a product.

Five of the 21 polymorphisms observed were variable-length repeats of one or two bases. Similar variations in the length of GT or C repeats in HSV sequences due to DNA polymerase error have been reported elsewhere (9, 10). These polymerase errors are possible at several points in the pathogenesis of HSV-2. Mutations may have occurred during viral replication in the recurrent lesion of partner A or the primary lesion of partner B. Alternatively, a variant strain may have colonized the ganglia of partner A in the intervening time between sample collection from partner A and partner B, which was 31 days for pair 8 and 51 days for pair 12.

We find it unlikely that HSV-2 sequence variation was introduced by the methods used in our study. First, even though mutations in DNA sequences can be introduced by the polymerases used in PCR and sequencing, all polymorphisms were confirmed by second-strand sequencing, and nucleotide differences between partners were confirmed by a second PCR and sequencing using the DNA harvested from culture. Second, we found the sequences to be stable when DNA samples taken directly from the lesion swab were compared to samples obtained after 2 passages in cell culture. There are no data to suggest that genetic drift in 2 passages in mink lung cells would be any more or less likely than drift in passages in HDF cells. Sequence drift during culture has been reported in noncoding hypervariable regions of HSV-1 (14); however, we found no such drift in our study.

We expected to observe increased variability between individuals in noncoding regions due to our hypothesized absence of purifying selection in these regions. Our study confirmed that noncoding regions exhibit a greater level of variability than similar-length segments of glycoprotein-coding regions. It is possible that, during HSV-2 replication, DNA polymerase errors in noncoding regions may not decrease the fitness of the virus. This is in contrast to the coding regions, which evolve with functional restraints and therefore are expected to be more conserved. The increased variability of noncoding regions facilitates the detection of variation with a much smaller, and thus more easily obtainable, amplicon. Overall, noncoding regions may be preferable for distinguishing HSV-2 strains at an individual level.

While we believe our major conclusions are firm, we do recognize that we sampled a small portion of the HSV-2 genome. While our source partners did not acquire HSV-2 exclusively in Seattle, we did analyze only a limited number of persons. It is possible that sequencing of HSV strains from a larger geographical area may show greater variability. However, it is noteworthy that HG52 differed from the consensus sequence only at three sites, even though its origin is geographically distant from that of our samples (7). In addition, the amplified segments of gC, gB, and gG were chosen by ease of primer selection and amplification. These factors may have influenced the selection of a more conserved region. Consequently, the variation we observed in those segments may not be indicative of the level of variation present across each entire coding region, and further research is needed to determine to what extent these polymorphisms, as well as others, can be used to define and compare HSV-2 strains across individuals and populations. Specific attention must be given to sequencing efforts that analyze large numbers of individuals to refine the use of HSV-2 sequencing for molecular epidemiology. Furthermore, DNA sequencing of consecutive isolates from an individual can confirm whether an HSV-2 infection in a single individual is limited to a single strain or whether more than one strain has colonized the ganglia. While existing literature suggests that most infections occur with a single strain, infection with more than a single strain has been described using RFLP (5, 19, 21).

In conclusion, we found that DNA sequencing was more powerful than both RFLP and HMA in distinguishing HSV-2 strains. The sequences obtained from noncoding regions provided the most information on differences between unrelated
samples. Our results suggest that a multilocus assay based on several DNA sequences has the potential to be an informative tool for identifying specific strains in an HSV-2 transmission event.

ACKNOWLEDGMENTS
This work was supported by NIH grant AI-30731. We acknowledge Lizette Embuscado for her technical assistance with the RFLP analysis.

REFERENCES