Genomic Sequence Analysis of Epstein-Barr Virus Strain GD1 from a Nasopharyngeal Carcinoma Patient†

Mu-Sheng Zeng,1,2 Da-Jiang Li,1,2 Qing-Lun Liu,1,2 Li-Bing Song,1,2 Man-Zhi Li,1,2 Ru-Hua Zhang,1,2 Xing-Juan Yu,1,2 Hui-Min Wang,1,2 Ingemar Ernberg,3 and Yi-Xin Zeng1,2*

State Key Laboratory of Oncology in Southern China1 and Departments of Experimental Research,2 Sun Yat-sen University Cancer Center, Guangzhou 510060, People’s Republic of China, and Microbiology and Tumor Biology Center, Karolinska Institutet, Box 280, S-171 77, Stockholm, Sweden3

Received 2 June 2005/Accepted 22 September 2005

To date, the only entire Epstein-Barr virus (EBV) genomic sequence available in the database is the prototype B95.8, which was derived from an individual with infectious mononucleosis. A causative link between EBV and nasopharyngeal carcinoma (NPC), a disease with a distinctly high incidence in southern China, has been widely investigated. However, no full-length analysis of any substrain of EBV from this area has been reported. In this study, we analyzed the entire genomic sequence of an EBV strain from a patient with NPC in Guangdong, China. This EBV strain was termed GD1 (Guangdong strain 1), and the full-length sequence of GD1 was submitted to the GenBank database. The assigned accession number isAY961628. The entire GD1 sequence is 171,656 bp in length, with 59.5% G+C content and 40.5% A+T content. We detected many sequence variations in GD1 compared to prototypical strain B95.8, including 43 deletion sites, 44 insertion sites, and 1,413 point mutations. Furthermore, we evaluated the frequency of some of these GD1 mutations in Cantonese NPC patients and found them to be highly prevalent. These findings suggest that GD1 is highly representative of the EBV strains isolated from NPC patients in Guangdong, China, an area with the highest incidence of NPC in the world. Furthermore, these findings provide the second full-length sequence analysis of any EBV strain as well as the first full-length sequence analysis of an NPC-derived EBV strain.

Epstein-Barr virus (EBV) is a ubiquitous human gammaherpesvirus that infects more than 90% of the population worldwide. It has been implicated in the development of several malignancies of epithelial or lymphoid origin, including undifferentiated nasopharyngeal carcinoma (NPC) (27), rare cases of gastric carcinoma (23), endemic Burkitt’s lymphoma (60), Hodgkin’s disease (54), B- and T-cell lymphoma (37), and B-cell lymphoproliferations in immunocompromised patients (20).

In order to obtain large amounts of virus for genetic and functional analyses, various EBV strains, including B95.8, P3HR-1, Raji, AG876, W91, Jijoye, Akata, and M-ABA, have been established. The prototype EBV strain is derived from the B95.8 cell line, which was established by infecting marmoset B cells with EBV from the 883L cell line. The 883L cell line was obtained by culture of lymphocytes from an individual with infectious mononucleosis (36, 48). 883L virus conserved the EcoRI C fragment which was lost in the B95.8 virus. The Jijoye cell line was derived from an African boy with Burkitt’s lymphoma (39). P3HR-1 is a subline of the P3 (Jijoye) Burkitt’s lymphoma cell line, which constantly produces infectious EBV (21); in contrast, the Raji line is non-virus-producing (17, 39). The P3HR-1, Raji, AG876, W91, and Jijoye lines are all African Burkitt’s lymphoma isolates, whereas the Akata is a cell line derived from a Japanese patient with Burkitt’s lymphoma (52).

The M-ABA line was initially derived from malignant tumor cells of an Algerian individual with nasopharyngeal carcinoma. The M-ABA line was ultimately obtained by cocultivating these cells grown in a nude-mouse-grown tumor with human mononuclear cells from EBV-seronegative adults. An improvement in virus production was therefore obtained by passaging the transforming virus into marmoset lymphocytes, resulting in the M-ABA line, which provides the first abundant source of NPC-derived infectious EBV for comparative studies (12). However, to date, the prototype B95.8 is the only EBV line with a complete genomic sequence available in the database (4); thus, it has been extensively mapped for transcripts, promoters, open reading frames, and other structural elements (GenBank accession no. V01555). Furthermore, although B95.8 has been widely used in studies of EBV, the DNA sequences and restriction maps of other EBV strains are only partly known, and no genomic sequence of EBV from NPC has been completed.

It has been suggested that EBV may play a role in the development of NPC, as no other type of tumor in humans is as consistently associated with EBV as NPC (27). Despite the fact that EBV infection is ubiquitous, the incidence of NPC presents a remarkable geographic pattern, as it is nearly 100 times more frequent in Southeast Asia, North Africa, and Alaska than in the rest of the world. Although various cultural or environmental cofactors have been suggested to explain this unique geographic distribution (41), the possibility of the existence of highly tumorigenic EBV strains in these areas cannot be excluded. If this hypothesis is true, one would expect to find these strains in EBV-associated tumors within the NPC-affected regions. Although the existence of variations in EBV
genes within different geographic regions has been extensively investigated (7, 14, 16, 46), no disease-associated EBV subtype has been defined thus far. Therefore, it is particularly important to analyze the entire genomic sequence of an NPC-derived EBV strain from an affected region, such as southern China, which may allow researchers to identify a potential NPC-related specific EBV strain from the overall sequence rather than from the sequence of only a few known genes.

In the current study, we infected umbilical cord blood mononuclear cells with EBV obtained from a Cantonese patient with nasopharyngeal carcinoma. From this, we created a lymphoblastoid cell line (LCL) that we subsequently named GD1 (GenBank accession no. AY961628). The genome of GD1 EBV was sequenced and was compared to B95.8 (V01555). In so doing, mutations, including deletions, insertions, and point mutations, were identified in GD1. The existence of these mutations suggests that GD1 is a unique strain that importantly differs from B95.8 and, notably, is widespread in Guangdong, China.

**Materials and Methods**

**Subject.** Saliva was collected in 1998 from a 39-year-old male Cantonese patient with nasopharyngeal carcinoma prior to treatment at the Sun Yat-sen University Cancer Center, Guangzhou, China. The patient’s consent as well as approval from the Institute Research Ethics Committee were obtained for use of these clinical materials for research purposes. Histopathologically, the primary tumor was diagnosed as a nonkeratinizing carcinoma (WHO type II). The clinical stage was determined to be T4N1M0 (IVA). The serological titrations of immunoglobulin A antibodies against the EBV viral capsid antigen and the diffuse component of early antigen are 1:640 and 1:20, respectively. The patient was radioreacted treatment and remains disease free.

**Establishment of a lymphoblastoid cell line.** Umbilical cord blood mononuclear cells were isolated on Ficoll-Hypaque density gradients, and 10⁶ cells were then infected for 2 h with EBV in 1 ml of the collected saliva. After infection, the cells were washed with phosphate-buffered saline and then cultured in RPMI 1640 medium supplemented with 10% fetal calf serum ( Gibco, Grand Island, NY) in the presence of 50 ng/ml cyclosporine (Sandoz, Basel, Switzerland). The cultures were fed weekly with fresh medium. Small colonies began to appear within 20 to 50 days. Individual colonies were then expanded and maintained for DNA extraction.

**PCR amplification of EBV fragment.** DNA from the LCL was extracted by phenol-chloroform. The PCR primers used to sequence the EBV genome are shown in Table S1 in the supplemental material. PCR was performed using 0.5 μg of DNA, 0.5 μM concentrations of each primer, 200 μM concentrations of each dideoxynucleoside triphosphate, 10 mM Tris-HCl (pH 8.8), 50 mM KCL, 1.5 mM MgCl₂, 0.2 mg/ml bovine serum albumin, and 2.5 U of Taq polymerase (TaKaRa, Dalian, China) in a total volume of 100 μl. Thirty cycles of denaturation (94°C for 45 s), annealing (55°C for 45 s), and extension (72°C for 1 to 10 min) were carried out in an automated thermal cycler (Peltier Thermal Cycler). The amplified mixture was electrophoresed on agarose gels, and the specific bands were cut and purified.

**Sequence analysis.** The purified PCR products were sent to the Beijing Genomics Institute, Chinese Academy of Sciences, and processed by shotgun sequencing, as described previously (43). Several small fragments were directly sequenced using the PCR primers. Every nucleotide was covered by at least two identical readings. However, the numbers of long gaps were linked by direct PCR sequencing. The entire GD1 sequence is 171,656 nucleotides (nt) in length, with G+C content comprising 59.5% of the total and A+T comprising 40.5% of the total. Due to the similarity of GD1 to the B95.8 sequence in the EBNA2 fragment, the deletions in the EBNA3B (nt 84883 to 85,007) and EBNA3C (nt 87489 to 87641) fragments, as well as the absence of a deletion in the EBNA3A (nt 81173 to 81447) fragment, GD1 was classified as type A according to the guidelines established in previous reports (2, 42, 44) and as a China 1 strain based on the standards described by Edwards et al. (15). The absence of a BamHI site between BamHI W1 and I1 indicated that GD1 is a “c” variant (41). However, GD1 is not characterized as an “f” variant, which results from an additional BamHI cut site at the left end of the genome (33). Furthermore, GD1 retained the XhoI site at the right end of the genome, which has been reported to be lost in some EBV genomes from NPC samples (22). Variations, including nucleotide insertions, deletions, and point mutations, were detected and compared with the genomic sequence of B95.8.

**Nucleotide deletions.** Compared with the entire B95.8 sequence, GD1 contained deletions in 43 sites (Table 1). Nineteen deletions were single-nucleotide deletions, and four were dinucleotide deletions. Among the single-nucleotide deletions, only one deletion was located in a predicted open reading frame (BPLF1) expressing an unnamed protein product, but the deletion was coupled with a single-nucleotide insertion three nucleotides away. As a result, two amino acid substitutions (GA->EG) are predicted to occur in BPLF1. Other deletions that resulted in no changes within the coding sequence (CDS) existed in promoter, poly(A) downstream, or repeat regions. Twenty deletions...
involved more than two nucleotides, and the number of deleted nucleotides was always three or a multiple of three, which resulted in no CDS frameshift, but 10 of these deletions were located in exons of various genes and resulted in an amino acid deletion. The other 10 deletions were located in promoter, poly(A) signal, or repeat regions.

A 30-bp deletion in the C terminus of LMP1 has also been observed in the GD1 strain, and this deletion has been sug-
gested to be associated with the oncogenic potential of LMP1 (31). In addition, it is interesting that the four fragment deletions in BLLF1 (gp350) are also detected in GD1; these deletions have previously been reported by Desgranges et al. (13).

These four deletions included two 21-bp, one 27-bp, and one 84-bp deletion, resulting in losses of 7, 9, and 28 amino acids, respectively, but no concomitant frameshift. Other two-fragment deletions were located in the genes EBNA3C (EBNA

<table>
<thead>
<tr>
<th>Nucleotide position in GD1</th>
<th>Inserted nucleotide(s)</th>
<th>Gene and function of insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>5812–5813</td>
<td>CC</td>
<td>BFLF1 reading frame</td>
</tr>
<tr>
<td>6391–6392</td>
<td>GT</td>
<td>Unknown function between promoter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and Pol III RNA EBER 1</td>
</tr>
<tr>
<td>7127</td>
<td>T</td>
<td>Pol III RNA EBER 2</td>
</tr>
<tr>
<td>7718–7751</td>
<td>GAGATTAGGATAGCATATGCTACCTAATCTCTAGTCCGGGCTGCTGAGTGGAGGA</td>
<td>Tandem repeat part of oriP</td>
</tr>
<tr>
<td>10834</td>
<td>C</td>
<td>Unknown function between poly(A) signal and promoter</td>
</tr>
<tr>
<td>14514</td>
<td>T</td>
<td>Exon W1 (also W66) part of leader protein (LP) gene</td>
</tr>
<tr>
<td>17591</td>
<td>T</td>
<td>3072 repeat 2 (BamHI W fragment)</td>
</tr>
<tr>
<td>20664</td>
<td>T</td>
<td>3072 repeat 3 (BamHI W fragment)</td>
</tr>
<tr>
<td>23737</td>
<td>T</td>
<td>Unknown function between promoter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exon W1</td>
</tr>
<tr>
<td>26810</td>
<td>T</td>
<td>3072 repeat 6 (BamHI W fragment)</td>
</tr>
<tr>
<td>29883</td>
<td>T</td>
<td>Downstream of promoter</td>
</tr>
<tr>
<td>32926</td>
<td>T</td>
<td>BFLF1 reading frame</td>
</tr>
<tr>
<td>43198–43200</td>
<td>GTT</td>
<td>Unknown function between poly(A) signal and BORF2 early reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>frame</td>
</tr>
<tr>
<td>57568–57618</td>
<td>GGAGCCCGGGCGGCTGCTGTGAGTGGAGGA</td>
<td>2 71-bp repeats</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EBNA3C; 3 39-bp repeats</td>
</tr>
<tr>
<td>58172–58201</td>
<td>AGGCCTGAGCTGCAAGGCGGAGGCCGAGAGAGGAGGA</td>
<td>BOLF1 reading frame</td>
</tr>
</tbody>
</table>
| 59383–59385               | GCC                    | Unknown function between poly(A) signal |}

These nucleotide insertions in the GD1 strain were compared to a previously published EBV sequence (AJ507799), since the fragment is deleted in B95.8.
TABLE 3. Types of point mutations in the GD1 strain compared to the B95.8 strain

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T&gt;C</td>
<td>229</td>
<td>58.59%</td>
</tr>
<tr>
<td>G&gt;C</td>
<td>193</td>
<td>13.66%</td>
</tr>
<tr>
<td>G&gt;A</td>
<td>206</td>
<td>14.58%</td>
</tr>
<tr>
<td>A&gt;G</td>
<td>200</td>
<td>14.15%</td>
</tr>
<tr>
<td>Transversions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A&gt;T</td>
<td>27</td>
<td>1.91%</td>
</tr>
<tr>
<td>T&gt;A</td>
<td>39</td>
<td>2.76%</td>
</tr>
<tr>
<td>A&gt;C</td>
<td>94</td>
<td>6.65%</td>
</tr>
<tr>
<td>C&gt;A</td>
<td>87</td>
<td>6.16%</td>
</tr>
<tr>
<td>T&gt;G</td>
<td>112</td>
<td>7.93%</td>
</tr>
<tr>
<td>G&gt;T</td>
<td>103</td>
<td>7.29%</td>
</tr>
<tr>
<td>C&gt;G</td>
<td>59</td>
<td>4.18%</td>
</tr>
<tr>
<td>G&gt;C</td>
<td>64</td>
<td>4.53%</td>
</tr>
<tr>
<td>Total</td>
<td>1,413</td>
<td>100</td>
</tr>
</tbody>
</table>

4B) and EBNA2, resulting in losses of 34 and 35 amino acids, respectively (Table 1).

**Nucleotide insertions.** We identified 44 nucleotide insertions in the GD1 strain compared to the B95.8 sequence (Table 2). Thirty-one insertions were single-nucleotide insertions, and two were dinucleotide insertions. The insertions, in general, existed in promoter downstream, poly(A) signal, or repeat regions and resulted in no CDS changes. Two exceptions to this included a single-nucleotide insertion in the BPLF1 fragment, which was coupled with a single-nucleotide deletion three nucleotides away, as mentioned above, and another single-nucleotide insertion in LF2. The dinucleotide insertion existed in the open reading frame of the predicted gene BFLF1 and resulted in a stop code which may eliminate the predicted reading frame. Another 11 insertions involved three or more nucleotides. One was located in the intron of BZLF1. Three insertions were located in promoter downstream, poly(A) signal, or repeat regions. The other seven were located in open reading frames of various genes, resulting in insertions of amino acids in the repeat region of EBNA3C, the predicted gene BPLF1, and BOLF1. Notably, there was an 11,830-bp insertion in GD1 strain between nucleotide positions 152012 and 152013 of B95.8, but this nearly 12-kb fragment also exists in the P3HR-1, Raji, W91, and M-ABA strains (6, 38).

**Nucleotide point mutations.** We identified numerous point mutations in GD1 compared to B95.8 (see Table S2 in the supplemental material). To summarize, the 1,413 point mutations included 856 single-nucleotide polymorphisms in coding regions, among which 335 were missense and 521 were synonymous. The remaining mutations were located in noncoding regions. All point mutations were distributed throughout the entire GD1 genome, and most genes were affected. The types of mutations in GD1 were also classified as transition and transversion, and there were more transition (828; 58.59%) than transversion (585; 41.41%) mutations (Table 3).

**GD1 is representative of EBV found in Cantonese NPC patients.** Because we detected such a large number of mutations in GD1 compared to B95.8, it was not clear whether GD1 accurately reflected the sequence variations in EBV in a larger population of Cantonese NPC patients. We then evaluated the extent to which GD1 is representative of EBV strains in Cantonese patients by measuring the frequency of various mutations identified in GD1 and comparing this to the frequency of these mutations in EBV from other Cantonese NPC biopsy specimens. As shown in Table 4, the 16 mutations in EBNA1 and the C terminus of LMP1 genes identified in GD1 were detected with high frequency (84 to ~100%) in 54 NPC biopsy specimens from Cantonese patients. At amino acid position 487 of EBNA1, GD1 contains a valine which is common in NPC patients (5, 19, 49). Nucleotide changes at position 168795 (T>G) resulted in an amino acid change (L->I) at amino acid position 25 in the N-terminal region of LMP1. This mutation has been used as a criterion for EBV typing in a previous study (51). The commonly observed 30-bp deletion at nt 90104 to 90130 in BLLF1 compared with B95.8 were detected in GD1 (Table 1). Considering the variations in the signature sequence in the C terminus of LMP1 (15), GD1 belongs to the China 1 strain. Nucleotide changes in comparison with B95.8 at positions 6372 (T>G), 6887 (G>A), 6889 (T>G), 6914 (A>G), 6947 (G>A), and 7126 (A>G) were found in GD1, indicating that GD1 shares the basic features of type A EBV. The first six of the above-mentioned nucleotides were located in the region between EBER1 and EBER2 CDS. The last nucleotide, 7126 (A>G), exists in the promoter downstream of EBER2 (3, 32). These results are consistent with our previous findings (58). Nucleotide substitutions at positions 159926 (G>T), 159998 (C>T), 160154 (C>A), and 160160 (G>C), resulting in amino acid changes in the BARF0 open reading frame, were observed in GD1 in comparison with B95.8, and those variations have been found to be highly frequent among Cantonese NPC patients (50). Taken together, these results suggest that GD1 is highly representative of the EBV strains found in Cantonese NPC patients.

**DISCUSSION**

Because of the distinct geographical and ethnic distribution of NPC, both the host genomic background and the variability of EBV strains have been suggested to influence the carcino-
genesis of NPC. We previously localized an NPC susceptibility locus to chromosome 4p12–p15 by an entire genomic scan linkage analysis (18). To determine the existence of sequence variations within EBV strains in a region with a high incidence of NPC, we started by establishing an LCL from umbilical cord blood mononuclear cells transformed by saliva virus from a Cantonese NPC patient. In order to determine whether the EBV strain from this LCL is representative of the EBV commonly found in this area, the 30-bp deletion in the C terminus of the LMP1 gene, which has been reported to be prevalent in Cantonese NPC samples, was analyzed and was found to exist in this strain. The strain was named GD1 and expanded for further analysis. Initially, we attempted to induce and purify virus directly from the cells. Although the purified virus particles could be detected by electron microscopy, the amount of virus was not sufficient to establish a viral DNA library for direct sequencing. The development of long-distance PCR techniques has provided an alternative method to establish the viral DNA library. This strategy, although labor-intensive, was successful.

GD1 greatly resembled the prototypical EBV strain B95.8. The entire sequence of GD1 is 171,656 bp, only 625 bp shorter than B95.8, with 59.5% G+C content and 40.5% A+T content, similar to the content found in B95.8 (59.94% and 40.06%, respectively). Both B95.8 and GD1 are type A strains. In addition, GD1, like B95.8, also contains a preserved XhoI site at the right end of the genome and has no “f” variant. GD1 was annotated according to B95.8 (V01555) and shared of the majority of the open reading frames and promoters within B95.8. In general, GD1 and B95.8 were highly homologous; however, the 43 deletions, 44 insertions, and 1,413 point mutations in GD1 suggest that GD1 is a divergent EBV strain. Among the identified EBV strains, the B95.8 viral DNA is unique in its 12-kb deletion in the EcoRI C fragment (40). Similar to P3HR-1, Raji, W91, and AG876 (35), GD1 has an additional 12 kb of DNA in the EcoRI C fragment.

Because EBV infection is associated with the development of a variety of diseases, the possibility of disease-specific subtypes of EBV has been widely studied. The early studies indicated that the EBV strains isolated from different regions of the world and from patients with different EBV-associated diseases were closely related (6). However, EBV isolates now can be classified into two distinct families (types A and B) by examination of EBNA2, EBNA3, and EBERs. These two types, also known as EBV-1 and EBV-2, exhibit different biological functions, as demonstrated by in vitro transformation assay (42), although both type A and type B EBV similarly influence the pathogenesis of Burkitt’s lymphoma (55). As determined by the sequences of EBNA2, EBNA3, and EBERs, GD1 belongs to type A EBV. Although it has been reported that type A EBV is more prevalent in Chinese NPC, it is unlikely to be a disease-specific subtype of EBV (6).

By various independent sorting methods of the two types, EBV can be further classified into different strains. Restriction enzyme polymorphisms have been used for the initial characterization. A prevalent Chinese strain that lacks the BamHI restriction site between the W1 and II fragments gains an extra site within the BamHI F fragment (1, 59), and lacks an XhoI restriction site within the N-terminal part of the LMP1 gene (10, 25) has been described, but only the lack of the BamHI restriction site between the W1 and II fragments is present in the GD1 strain. Therefore, GD1 is clarified as “F” type. Although the virus detected in saliva could be different from the virus present in the NPC tumors, which has been observed for the “F” variant as well as for the H-polymorphism in previous studies (33, 47), both PCR-restriction fragment length polymorphism and PCR-sequencing analysis detected the virus from the paraffin-embedded NPC tumor of the patient in this study as the “F” type (data not shown).

Further studies have examined the sequence variation in the LMP1 gene (26, 28–30), which is considered to be the main EBV oncogene. Most LMP1 genes derived from Chinese NPC biopsy specimens are marked by the 30-bp deletion (nt 168266 to 168295) in the carboxyl terminus, a loss of the XhoI site in the amino terminus, and multiple base substitutions in the coding region (9, 10, 22, 25). Although the 30-bp deletion in LMP1 was detected in 100% of 48 NPC biopsy specimens in Taiwan (8), in 34 of 37 NPC biopsy specimens in Hong Kong (11), and in 16 of 21 NPC biopsy specimens in Guangxi and Shanghai (51). Zhang et al. (56) reported that the 30-bp deletion in LMP1 represents a geographical- or race-associated polymorphism rather than an NPC disease phenotype-associated polymorphism. Recently, the consistent sequence variation in LMP1 has been used to distinguish the various EBV strains, which were termed strains China 1, China 2, Med, China3, Alaskan, NC, and B95.8. In a study with a large number of samples, LMP1 with the 30-bp deletion and China 1 were found in 86% of 187 NPC biopsy specimens in Asia (15). According to the sequence of amino acids at positions 229 (Ser), 334 (Gln to Arg), and 344 (deleted), GD1 belongs to the China 1 strain, which is the predominant strain in China (15).

Another latent-expressing gene used for EBV strain analysis is EBNA1, which is the only gene consistently expressed in almost all EBV-infected cells studied to date and is essential for the persistence and replication of the virus in latently infected cells. Based on the 487 amino acids in EBNA1, the EBV strains can be classified into five subtypes, two prototypes (P-Ala and P-Thr), and three variant subtypes (V-Pro, V-Leu, and V-Val). Snudden et al. (49) reported that V-Val was the only subtype of EBNA1 present in seven NPC biopsy specimens from Hong Kong. Subsequently, Gutierrez et al. (19) observed that V-Val was detected in 9 of 13 NPC biopsy specimens from Hong Kong and France. Meanwhile, multiple subtypes of EBNA1 have been detected in throat washings (TWs) and peripheral blood lymphocytes from 21 healthy donors. In TWs, V-Val was identified in only 25% of healthy donors, but mixed infection of V-Val with other subtypes of EBNA1 was observed. Recently, Wang et al. (53) reported that V-Val was detected in 13 of 13 NPC tissues in Taiwan. However, Sandvej et al. (45) observed that the subtypes of EBNA1 in NPC had no tumor-specific expression. Although V-Val was found in three of four NPC biopsy specimens from Beijing in northern China, a similar prevalence of EBNA1 occurred in TWs in the background population. Likewise, P-Thr was reported to be prevalent in both Danish NPC patients and healthy populations. Zhang et al. (57) reported that V-Val was the only subtype detected in NPC tissue and suggested that a substrate of EBV with the V-Val subtype of EBNA1 preferentially infected NPC patients. Mai et al. (34) also reported that 91.67% (33/36) of the EBNA1 subtypes from NPC biopsy specimens were single
V-Val subtypes, and the remaining three patients were coinfected with V-Val and P-Ala. These findings suggest that V-Val might be a subtype that adapts particularly well to the nasopharyngeal epithelium or that mutations in V-Val endow the virus with increased oncogenic potential (34, 57). GD1 is a V-Val variant, and previously we also found this type to be the major subtype in NPC tissues (34, 57). This result provides another piece of evidence that GD1 well represents EBV isolated from NPC patients.

In addition to LMP1 and EBNA1, mutations in other latently expressed EBV genes, such as BARF0 and EBERs, have also been identified in NPC samples (3, 32, 50, 58). Janse van Rensburg et al. (24) reported that EBV mutants have features of type 2 EBV strains, but Zhang et al. found EBV from NPC tissue to exhibit the basic characteristics of type 1 EBV strains and also detected eight base pair substitutions and two 1-bp deletions in EBER2. The sequences of EBERs from normal nasopharynxes in Guangdong and from throat washings in Haerbin were identical to those from NPC tissue in Guangdong (58). GD1 contains mutations in EBERs, consistent with Zhang’s report (57). Song et al. (50) initially reported four mutations in BARF0 in 100% of NPC biopsy specimens, and those mutations were also detected in GD1 in this study. Therefore, based on the high frequency of the mutations in the various major genes of GD1 detected in Cantonese NPC patients, GD1 may represent an EBV strain prevalent in NPC from Guangdong, China.

In summary, our data provide the second full-length sequence analysis of an EBV strain as well as the first full-length sequence analysis of an NPC-derived EBV strain. Our results indicate that GD1 is a unique EBV strain that differs from the prototypical B95.8 strain but is prevalent in Cantonese NPC patients.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Science and Technology of China (2002BA711A03 and 2004CB518801), Guangdong Province Science and Technology Project (2004A30801002), and the National Natural Science Foundation of China (30470666 and 30440001).

REFERENCES


