SHORT COMMUNICATION

The entire nucleotide sequence of two hepatitis G virus isolates belonging to a novel genotype: isolation in Myanmar and Vietnam

Hideo Naito,1, 2 Shigeki Hayashi2 and Kenji Abe1

1 Department of Pathology, National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640, Japan
2 Division of Gastroenterology, International Medical Center of Japan, Tokyo, Japan

A novel genotype of hepatitis G virus (HGV) was recently identified in sera of subjects from countries in South-East Asia. These isolates were recovered from serum of Myanmarese (designated HGV-MY14) and Vietnamese (designated HGV-VT48) subjects, respectively. To characterize the viral genome in more detail, the full-length nucleotide sequence of the two different HGV isolates belonging to the novel genotype was cloned. Both HGV isolates were composed of 9228 nt and had a single open reading frame spanning 8529 nt and encoding 2843 aa residues. The isolates differed from previously reported HGV/GBV-C isolates types 1 to 3 by 13–15% (nucleotide sequence) and 4–6% (amino acid sequence). The putative core region of both isolates was not clearly identifiable as it consisted of only 16 aa residues. Based on phylogenetic analysis of full-length genome sequences and 5′-UTR sequences, HGV-MY14 and HGV-VT48 isolates can be classified as a novel genotype, designated type 4.

In 1996, novel RNA viruses were identified in the sera of patients with liver disease in two different American groups. These agents have been named hepatitis G virus (HGV) and hepatitis GB virus type C (GBV-C), respectively (Linnen et al., 1996; Leary et al., 1996). Molecular characterization of these two agents has shown them to be different isolates of the same virus. They are single-stranded RNA viruses approximately 9–10 kb long with high similarity (96%) at the amino acid level (Zuckerman, 1996). The agents also have characteristics of a flavivirus-like genome, e.g. like hepatitis C virus (HCV), and may represent a new genus in the family Flaviviridae, which includes flaviviruses and pestiviruses (Muerhoff et al., 1995). It is also known that HGV/GBV-C isolates are genetically heterogeneous; three major genotypes, types 1 to 3, have been identified so far (Muerhoff et al., 1996, 1997; Orito et al., 1996; Takahashi et al., 1997a, b; Viazov et al., 1997; Saito et al., 1998). In addition to this classification, we recently identified a novel genotype of HGV in South-East Asian countries, including Myanmar and Vietnam (Naito et al., 1999). Investigation of the sequence diversity of different virus isolates is important because variants may differ in characteristics such as their patterns of serological reactivity, pathogenicity, virulence and response to therapy. In this study, we cloned the entire nucleotide sequence of two HGV isolates belonging to the novel genotype to clarify the molecular characterization of this viral genome.

Two different HGV isolates, recovered from serum specimens of a Myanmarese subject [29-year-old female, infected with HCV and human immunodeficiency virus (HIV)] and a Vietnamese subject [17-year-old female, infected with hepatitis B virus and HIV], respectively, were cloned and sequenced. These isolates were designated HGV-MY14 and HGV-VT48, respectively. The methods for RNA extraction from serum, amplification of HGV cDNA and cloning strategy have been described previously (Abe et al., 1997; Kaneko et al., 1998). The sequence of PCR primers is listed in Table 1. The terminal sequence of the 3′-UTR of the HGV genome was isolated by a rapid amplification of the cDNA ends (RACE) kit (5′/3′ RACE kit; Boehringer Mannheim). This method is based on a ‘one-sided’ PCR procedure. For the terminal sequence of the 3′-UTR, a poly(A) tail, which is lacking in HGV/GBV-C, was added to RNA extracted from serum by poly(A) polymerase (TaKaRa Biochemicals) before the RACE step. Briefly, RNA extracted by SepaGene RV-R (Sanko Junyaku) was dissolved in 10 µl RNase-free H2O and mixed with 40 µl buffer containing 50 mM Tris–HCl, pH 8.0, 10 mM MgCl2, 2.5 mM MnCl2, 250 mM NaCl, 1 mM DTT, 0.5 mg/ml BSA, 0.1 mM ATP and 1 U poly(A) polymerase. After incubation at 37 °C for 1 h, poly(A)-tailed RNA was ethanol-
Table 1. Primer sequences used for nested RT–PCR

<table>
<thead>
<tr>
<th>Clone</th>
<th>Primer</th>
<th>Nucleotide position*</th>
<th>Polarity</th>
<th>Sequence (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>HG1</td>
<td>139–158</td>
<td>+</td>
<td>GGT CGT AAA TCC CGG TCA CC</td>
</tr>
<tr>
<td></td>
<td>HG9R</td>
<td>957–976</td>
<td>-</td>
<td>CAA AAC TCA CTT TCC CAC TT</td>
</tr>
<tr>
<td></td>
<td>HG2</td>
<td>163–182</td>
<td>+</td>
<td>TAG CCA CTA TAG GTG GTG CT</td>
</tr>
<tr>
<td></td>
<td>HG8R</td>
<td>924–943</td>
<td>-</td>
<td>ACT GCA CAC GTC AGG TTA GG</td>
</tr>
<tr>
<td>b</td>
<td>HG9</td>
<td>762–777</td>
<td>+</td>
<td>GGT CGG CCT GCC AAG TCC GC</td>
</tr>
<tr>
<td></td>
<td>HG12R</td>
<td>2175–2197</td>
<td>-</td>
<td>CAA AGT CCA AGA GCA ACC AGC G</td>
</tr>
<tr>
<td></td>
<td>HG2</td>
<td>163–182</td>
<td>+</td>
<td>TAG CCA CTA TAG GTG GTG CT</td>
</tr>
<tr>
<td></td>
<td>HG8R</td>
<td>924–943</td>
<td>-</td>
<td>ACT GCA CAC GTC AGG TTA GG</td>
</tr>
<tr>
<td>c</td>
<td>HG9</td>
<td>762–777</td>
<td>+</td>
<td>GGT CGG CCT GCC AAG TCC GC</td>
</tr>
<tr>
<td></td>
<td>HG12R</td>
<td>2175–2197</td>
<td>-</td>
<td>CAA AGT CCA AGA GCA ACC AGC G</td>
</tr>
<tr>
<td></td>
<td>HG2</td>
<td>163–182</td>
<td>+</td>
<td>TAG CCA CTA TAG GTG GTG CT</td>
</tr>
<tr>
<td></td>
<td>HG8R</td>
<td>924–943</td>
<td>-</td>
<td>ACT GCA CAC GTC AGG TTA GG</td>
</tr>
<tr>
<td>d</td>
<td>HG6</td>
<td>2444–2464</td>
<td>-</td>
<td>GGT CCT CCA GTG GTG GAA GG</td>
</tr>
<tr>
<td></td>
<td>HG15R</td>
<td>4038–4057</td>
<td>+</td>
<td>GCC AGA AAC CCA CCT GTG CT</td>
</tr>
<tr>
<td></td>
<td>HG16R</td>
<td>3931–3950</td>
<td>+</td>
<td>CCC CAT GTA AGG GCC CAT GG</td>
</tr>
<tr>
<td>e</td>
<td>HG17</td>
<td>3848–3868</td>
<td>+</td>
<td>AAA GAG CAC TCG CCT CCC GTT</td>
</tr>
<tr>
<td></td>
<td>HG19R</td>
<td>5638–5656</td>
<td>+</td>
<td>GCG TTG CCT CCG ACG CCC A</td>
</tr>
<tr>
<td></td>
<td>HG18</td>
<td>3879–3896</td>
<td>-</td>
<td>AAC ATG GGG CAC AAG GTC</td>
</tr>
<tr>
<td></td>
<td>HG30R</td>
<td>4538–4562</td>
<td>-</td>
<td>GTA ATG GTG GGA TCA AGG TGC ACC</td>
</tr>
<tr>
<td>f</td>
<td>HG17</td>
<td>3788–3868</td>
<td>+</td>
<td>AAA GAG CAC TCG CCT CCC GTT</td>
</tr>
<tr>
<td></td>
<td>HG19R</td>
<td>5638–5656</td>
<td>-</td>
<td>GCG TTG CCT CCG ACG CCC A</td>
</tr>
<tr>
<td></td>
<td>HG4</td>
<td>4538–4563</td>
<td>+</td>
<td>GGT GAC CAC TGG TCC CAC CAT TAC C</td>
</tr>
<tr>
<td>g</td>
<td>HG4</td>
<td>4538–4563</td>
<td>+</td>
<td>GGT GAC CAC TGG TCC CAC CAT TAC C</td>
</tr>
<tr>
<td></td>
<td>HG7</td>
<td>4773–4797</td>
<td>+</td>
<td>TAC GAC GAC TGG TCT CCT TAC ACC GCC A</td>
</tr>
<tr>
<td></td>
<td>HG19R</td>
<td>5638–5656</td>
<td>+</td>
<td>GCG TTG CCT CCG ACG CCC A</td>
</tr>
<tr>
<td>h</td>
<td>HG20</td>
<td>5322–5346</td>
<td>+</td>
<td>GCC AAG ACA GTG ACA GGT GCG GTG G</td>
</tr>
<tr>
<td></td>
<td>HG20R</td>
<td>7102–7121</td>
<td>-</td>
<td>GGA AAG AGC CAC GTT GAA GA</td>
</tr>
<tr>
<td></td>
<td>HG19</td>
<td>5565–5584</td>
<td>-</td>
<td>GCT GCC GGC GTG GCC TA</td>
</tr>
<tr>
<td></td>
<td>HG7</td>
<td>6910–6934</td>
<td>-</td>
<td>ACA GGC ATC CTC CGG CTA CTA CCA C</td>
</tr>
<tr>
<td></td>
<td>HG5</td>
<td>4773–4797</td>
<td>+</td>
<td>TAC GAC GAC TGG TCT CCT TAC ACC GCC A</td>
</tr>
<tr>
<td></td>
<td>HG19R</td>
<td>5638–5656</td>
<td>+</td>
<td>GCG TTG CCT CCG ACG CCC A</td>
</tr>
<tr>
<td>i</td>
<td>HG7</td>
<td>6780–6804</td>
<td>+</td>
<td>GAA GAC TCA GAA GTG ACT GAG GCC G</td>
</tr>
<tr>
<td></td>
<td>HG5R</td>
<td>8705–8729</td>
<td>-</td>
<td>GCC CAC GAT GAT GTG AGG CAG TTT G</td>
</tr>
<tr>
<td>j</td>
<td>HG21</td>
<td>6882–6901</td>
<td>+</td>
<td>ATG GAG GAT TGG AGT ACA CC</td>
</tr>
<tr>
<td></td>
<td>HG21R</td>
<td>8431–8449</td>
<td>-</td>
<td>AGC CAA GTG GAG CAG AAG G</td>
</tr>
<tr>
<td>k†</td>
<td>HG29</td>
<td>8708–8728</td>
<td>+</td>
<td>ACT GCC TAA CAT CAT CCT GG</td>
</tr>
<tr>
<td></td>
<td>HG32</td>
<td>9048–9067</td>
<td>+</td>
<td>TTA GCC GTG CTC ATC GTA GC</td>
</tr>
</tbody>
</table>

* Nucleotide positions are based on HGV PNF2161.
† A single-sided PCR method was used.

Precipitated and subjected to the 3’ RACE technique as directed by the manufacturer. Amplification conditions included pre-incubation at 95 °C, 10 min activation of AmpliTaq Gold DNA polymerase (Perkin Elmer), followed by 35 cycles of PCR (94 °C, 20 s; 60 °C, 20 s; 72 °C, 60 s; with a final extension for 10 min, 72 °C). The recovered PCR products were re-amplified with internal HGV-specific primers. PCR products were separated by 1–2% agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen). Recovered PCR products were subcloned using a pBluescript II SK(–) vector (Stratagene) via the EcoRV site. Alternatively, purified PCR products were subjected to direct sequencing from both directions using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer). Sequences of amplified cDNA were determined using an ABI model 373A sequencer (Applied Biosystems).
Table 2. Nucleotide and amino acid identities (%) within the entire genome and subgenomic regions of HGV-VT48 for six reported HGV/GBV-C and HGV-MY14

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>Nucleotide (amino acid)</th>
<th>HGV/GBV-C genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Type 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GBV-C</td>
</tr>
<tr>
<td>Entire</td>
<td>9103–9395 (2842–2933)</td>
<td>86 (95)</td>
</tr>
<tr>
<td>5′-UTR</td>
<td>275–552</td>
<td>84</td>
</tr>
<tr>
<td>Core</td>
<td>48–321 (16–107)</td>
<td>92 (94)</td>
</tr>
<tr>
<td>E1</td>
<td>564 (188)</td>
<td>85 (96)</td>
</tr>
<tr>
<td>E2</td>
<td>1161 (387)</td>
<td>85 (92)</td>
</tr>
<tr>
<td>NS2</td>
<td>843 (281)</td>
<td>85 (95)</td>
</tr>
<tr>
<td>NS3</td>
<td>2031 (677)</td>
<td>86 (98)</td>
</tr>
<tr>
<td>NS4</td>
<td>945 (315)</td>
<td>87 (97)</td>
</tr>
<tr>
<td>NS5a</td>
<td>1245 (415)</td>
<td>89 (97)</td>
</tr>
<tr>
<td>NS5b</td>
<td>1692–1695 (564–565)</td>
<td>88 (94)</td>
</tr>
<tr>
<td>3′-UTR</td>
<td>61–312</td>
<td>72</td>
</tr>
</tbody>
</table>

Nucleotide sequences were multiple-aligned using CLUSTAL W version 1.4. The distance matrix of the nucleotide substitutions of each clone was estimated by the eight-parameter method (Rzhetsky, 1995) and phylogenetic trees were constructed from the matrix by the neighbour-joining method (Saitou & Nei, 1987). These procedures were computed using Phylo win version 1.2 (Galtier et al., 1996) on a DEC alpha 2000 server, and the trees were drawn by TreeView version 1.5 (Page, 1996). The reliability and topology of each tree branch was tested by bootstrap analysis (Billis & Bull, 1993) of the data of 1000 bootstrap resamplings of the columns in the full-length genome sequence alignment. Twenty-three full-length genome sequences and five nearly full-length genome sequences of HGV/GBV-C isolates obtained from databases were compared to the sequences of the isolates in the present study. The isolate names, accession numbers and references of the reported sequences are as follows: HGV PNF2161 and HGV R10291 (both USA isolates), U44402 and U45966, respectively (Linnen et al., 1996); GBV-C (West Africa), U36380 (Leary et al., 1996); HGV GA128 (Ghana), AB013500 (Saito et al., 1999); GBV-C/EA (East Africa), U63715 (Erker et al., 1996); HGV C964 (China), U75356 (Zhou et al., 1996); HGV CN (China), U94695 (Wang et al., 1997); HGV IM71 (Japan), AB008342 (Konomi et al., 1999); HGV Bl230 (Bolivia), AB003291 and AB003293, respectively (Takahashi et al., 1997b); K606, K1737, K1741, K1789, K1916, K2141, K1668, K1775 and K3732 (all Japan), D87708, D87709, D87710, D87711, D87712, D87713, D87714, D87715 and AB008335, respectively (Katayama et al., 1998).

The full-length genome sequences of HGV-MY14 and HGV-VT48 were obtained from serum samples from Myanmarese and Vietnamese subjects, respectively. Both isolates were composed of 9228 nt and contained a long open reading frame spanning 8529 nt and encoding 2843 aa flanked by putative 5′-UTR (nt 1–389 for MY14 and nt 1–387 for VT48) and 3′-UTR (nt 8919–9228 for MY14 and nt 8917–9228 for VT48) without a poly(U) stretch or poly(A) tail. Based on the predicted cleavage sites indicated by Mushahwar and colleagues (Muerhoff et al., 1995; Leary et al., 1996), we obtained the size of each region of the polyprotein in the isolate as follows: core fl 48 nt, 16 aa; E1 fl 564 nt, 188 aa; E2 fl 1161 nt, 387 aa; NS2 fl 843 nt, 281 aa; NS3 fl 2031 nt, 677 aa; NS4 fl 945 nt, 315 aa; NS5a fl 1245 nt, 415 aa; NS5b fl 1692 nt, 564 aa. The genomic organization was similar to that of previously reported isolates. As in prototypes of HGV and GBV-C isolates, the putative core region in MY14 and VT48 isolates could not be clearly defined as it consisted of only 16 aa. As shown in Table 2, when compared to other previously reported HGV/GBV-C isolates with a full-length genome sequence, HGV-VT48 showed overall identities of 85–87% at the nucleotide level and...
94–96% at the amino acid level, thereby indicating that they were the same virus. Among the coding regions, there was a tendency for the putative core region to show the highest similarity in nucleotide sequence with an identity of 88–96%. In contrast, lower similarities were observed in the E2 region. The 5′- and 3′-UTR were also highly conserved among the HGV/GBV-C genomes. HGV/GBV-C isolates with fully sequenced genomes, including database-derived sequences, were grouped into four major genotypes by phylogenetic analysis (Fig. 1). A similar finding was observed by phylogenetic analysis of the partial sequence of the 5′-UTR. To extend this investigation, we have cloned the entire nucleotide sequence of the two HGV isolates (designated HGV-MY14 and HGV-VT48, respectively), which were recovered from serum samples of Myanmarese and Vietnamese subjects and compared the sequences with those of previously reported full-length isolates. Since the discovery of the HGV/GBV-C genome in human serum in 1996, twenty-three full-length isolates of HGV/GBV-C have been reported. Our results revealed that the sequences of HGV-MY14 and HGV-VT48 isolates had a high level of similarity to those of database-derived HGV/GBV-C isolates. Furthermore, they had an incomplete putative core protein consisting of only 16 aa residues and no poly(A) tail, as is the case for other reported isolates of HGV/GBV-C.

To obtain the terminal sequence of HGV-MY14 and HGV-VT48, we designed poly(A)-tailed HGV RNA using poly(A) polymerase for use in RACE. By this method, poly(A) should be added only at the 3′ end of the RNA. The obtained genome showed that there was no poly(A) tail in the 3′-UTR and this strongly suggested that HGV-MY14 and HGV-VT48 do not have a further extended sequence such as the HCV 3′X tail (Tanaka et al., 1995). The nucleotide sequence in the 3′-UTR was highly conserved among the different HGV/GBV-C isolates except for the GBV-C genome. In the present study, we confirmed that HGV/GBV-C can be classified into at least four different genotypes by phylogenetic analysis of the entire sequence. Recently, we developed a rapid genotyping system of HGV/GBV-C by PCR using the type-specific primers. Using this new assay system, we screened the genotypic distribution of HGV in the world and confirmed that HGV type 4 is rather common in the south-eastern parts of the Asian continent, including Vietnam, Myanmar and Cambodia (un-
published results). The biological function of the putative structural and non-structural regions of the HGV/GBV-C genome is still unknown, and whether or not protective epitopes are encoded in the E1 and E2 regions remains to be investigated. Future studies of the significance of the diversity among the sequences of different HGV isolates, corresponding to the different genotypes, and HGV pathogenesis will need to take these considerations into account. It is to be hoped that primers deduced from highly conserved sequences, identified by the comparison of the full-length genomic sequences of HGV/GBV-C, will increase the sensitivity in detecting viral RNA in clinical studies and epidemiological surveys.

In conclusion, we have cloned the entire nucleotide sequence of two different HGV genomes recovered from the serum of Myanmarese and Vietnamese subjects, respectively. Both isolates (HGV-MY14 and HGV-VT48) were classified as a novel genotype, named type 4. It seems likely that HGV genotype 4 is rather common in south-eastern parts of the Asian continent. Additional studies, including analyses of sequences from other geographic regions, are required for the further classification of HGV/GBV-C.

We are very grateful to Takeshi Kurata, Department of Pathology, National Institute of Infectious Diseases, for his continuous encouragement, and to Teiichiro Shino, AIDS Research Center, National Institute of Infectious Diseases, for his expert advice in phylogenetic analyses. We also thank Khin Maung Win, Department of Hepatology, Yangon General Hospital in Myanmar, and Mari Yamaguchi and Mitsugu Usui, Sanko Junyaku Co., Ltd, Tokyo, for their kindy cooperation during this study. This study was supported in part by ‘Grants-in-Aid’ for Science Research of the Ministry of Education, Science and Culture of Japan and the Ministry of Health and Welfare of Japan.

References


Received 6 July 1999; Accepted 22 September 1999