Chronic hepatitis delta virus infection with genotype IIb variant is correlated with progressive liver disease

Hideki Watanabe, Kazuyoshi Nagayama, Nobuyuki Enomoto, Ryoko Chinzei, Tsuyoshi Yamashiro, Namiki Izumi, Hiroshi Yatsuhashi, Tatsunori Nakano, Betty H. Robertson, Hiroki Nakasone, Hiroshi Sakugawa and Mamoru Watanabe

Correspondence Nobuyuki Enomoto nenomoto.gast@tmd.ac.jp

1Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo, Tokyo 113-8519, Japan
2Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan
3Institute for Clinical Research, World Health Organization Collaborating Center for Reference and Research on Viral Hepatitis, National Nagasaki Medical Center, Nagasaki, Japan
4Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, USA
5First Department of Internal Medicine, School of Medicine, University of the Ryukyus, Okinawa, Japan

We determined the sequence of the hepatitis delta virus (HDV) genome in 40 Japanese patients, most of whom were from the Miyako Islands, Okinawa, Japan. Consensus sequences from 33 HDV full genomes out of a total of 40 patients were determined by directly sequencing four partially overlapping PCR products. Phylogenetic tree analysis classified these 33 complete HDV genomes as HDV genotype I (two patients), genotype IIa (one patient) and genotype IIb (30 patients). Among the 30 genotype IIb patients, there were two clusters of genetic variants. One group consisted of six isolates showing significant homology with genotype IIb, previously reported from Taiwan. The other group consisted of 24 isolates, whose sequences formed a new genetic subgroup (genotype IIb-Miyako; IIb-M). When the genetic structures were compared in detail between IIb and IIb-M, characteristic variations were found in the C-terminal sequence of the large delta antigen-conferring packaging signal as well as the RNA editing site. Determination of subclasses of genotype IIb in a total of 37 patients, including seven HDV patients whose partial HDV sequence was determined, revealed eight patients with IIb and 29 patients with IIb-M. Although there was no significant difference in the clinical background or virological state of hepatitis B virus between these two groups, patients with genotype IIb-M showed greater progression of chronic hepatitis and cirrhosis than those with genotype IIb (P = 0.0009). These data indicate the existence of a genetic subgroup of HDV genotype IIb, which is associated with different clinical characteristics and which could be related to genetic variations in functionally important parts of the HDV genome.

INTRODUCTION

Hepatitis delta virus (HDV) is a defective virus that requires hepatitis B virus (HBV) surface antigen for virion assembly (Rizzetto et al., 1980) and infection and contains a negative single-stranded circular RNA genome of 1.7 kb (Wang et al., 1986; Makino et al., 1987). HDV is classified into three genotypes (I, II and III) based on genetic sequence analysis (Casey et al., 1993). HDV genotypes correlate with the clinical outcome of HDV infection. HDV genotype I, which is found worldwide, often causes aggressive hepatitis and is more frequently associated with liver cirrhosis (LC) and hepatocellular carcinoma (HCC) than genotype II, which is mainly isolated from East Asia and is generally related to milder diseases (Wu et al., 1995a). On the other hand, acute infection with HDV genotype III, which is isolated from the northern part of South America, is closely associated with fulminant hepatitis (Casey et al., 1993). However, information about the precise relationship between the genetic structure of HDV and the clinical characteristics within each
genotype is somewhat limited. Comparative analysis, in which genetic variations are correlated with clinical presentations in a population with homogeneous clinical backgrounds or virological states of HBV, can provide valuable information about HDV genetic structures that determine the severity of liver disease.

In Japan, chronic HDV infection is relatively rare (Tamura et al., 1993) but is endemic in the Miyako Islands of Okinawa, where the HDV genotype II is prevalent (Sakugawa et al., 1999). Although the route by which HDV is spread on this island is unclear, our previous studies demonstrated that the severity of liver disease was heterogeneous within this population, despite relatively uniform clinical backgrounds. Thus, a detailed analysis in which the HDV genomes of these patients are correlated with clinical profiles could provide a unique opportunity to define the critical genetic features of HDV that determine liver injury.

To delineate the features of HDV isolates in this area, we determined the sequence of the full-length HDV genome from a large group of patients with chronic HDV infection, the majority of whom were from the Miyako Islands. As a result, we identified a new genetic variant of HDV genotype IIb that was associated with more progressive disease. Subsequently, specific genetic differences among these HDV genotype IIb isolates were correlated with the clinical features in order to reveal the variations in the HDV genome responsible for the progression of liver disease.

**METHODS**

**Patients.** A total of 40 patients were enrolled in the study; 16 males and 24 females ranging in age from 23 to 83 years old. Of the 40 patients, 37 were from the Miyako Islands of Okinawa, where HDV infection is endemic, two were from Nagasaki and one was from Tokyo, Japan. All patients showed positive serum anti-HD and HBsAg, including three asymptomatic carriers (ASC) who had consistently normal alanine aminotransferase (ALT) levels at least bimonthly for more than 2 years, 23 patients with chronic hepatitis (CH) with abnormal ALT levels and 13 patients with LC. The diagnosis of LC was based on clinical findings, such as oesophageal varices or ascites, with histological and/or radiological findings consistent with LC. The abdominal ultrasound findings for all ASC patients appeared normal without any evidence of liver disease. All subjects were hepatitis B e antigen (HBeAg) negative. To exclude other factors contributing to ALT elevation, the following subjects were excluded from the present study: subjects who were positive for antibody to hepatitis C virus (anti-HCV) or antinuclear antibodies, those with fatty liver on ultrasound examinations and those with a history of excess alcohol intake or hepatotoxic drugs. Written informed consent was obtained from patients in this study. The HBV DNA levels in the sera were quantified with a commercial kit (DNA probe Chugai-HBV; Chugai Diagnostics, Tokyo, Japan) using a transcription-mediated amplification assay (Kamisango et al., 1999). The detection range of this assay was from $5 \times 10^3$ to $5 \times 10^6$ copies ml$^{-1}$. HDV genotype was determined using the PCR-RFLP method (Mizokami et al., 1999).

**Sequencing of HDV.** The full-length HDV genome was sequenced in 33 patients. In the other seven patients, the partial genetic sequence encoding the delta antigen (HDAg) was determined. Extraction of RNA from 150 μl of serum by the acid guanidinium thiocyanate/phenol/chloroform method (Chomczynski & Sacchi, 1987) using ISOGEN (Wako, Osaka, Japan) and RT-PCR were performed as described previously (Enomoto et al., 1994). Four partially overlapping fragments were amplified by nested PCR using the primers shown in Table 1. These primers were designed and numbered based on HDV genotype II sequences in GenBank. PCR was initially performed with primers designed for HDV genotype II. If HDV cDNA was not amplified with these primers, PCR was performed with primers for HDV genotype I (primer sequences are available on request). Both strands of the PCR products were directly cycle sequenced with the PRISM dye termination kit (Applied Biosystems) and nested PCR primers.

**Sequence analysis.** The 33 HDV full genome sequences determined in this study were aligned with 22 complete nucleotide sequences of HDV retrieved from the international DNA databases (DDBJ/EMBL/GenBank). An initial alignment was made using the Clustal X 1.81 program (Thompson et al., 1997), followed by manual correction. Based on the alignment, a phylogenetic tree was constructed using the neighbour-joining method (Saitou & Nei, 1987), with genetic distances calculated using the Kimura two-parameter method using MEGA version 2.1 (http://www.megasoftware.net/). A maximum-likelihood tree was also constructed using PAUP 4.0 (D. L. Swofford, Sinauer Associates). To confirm the reliability of the phylogenetic trees, bootstrap resampling and reconstruction were carried out 1000 times for neighbour-joining trees and 100 times for maximum-likelihood trees. The nucleotide and amino acid identities were calculated using MEGA version 2.1. Phylogenetic

Table 1. The primer sets for the HDV genome sequence

<table>
<thead>
<tr>
<th>Outer primer</th>
<th>Inner primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1</td>
<td></td>
</tr>
<tr>
<td>5' CGAGGAGGAGCCGAGGGTGGGAGG (nt 16–39)</td>
<td>5' GCCAAAGAGTGGGGGAAATCTCG (nt 64–86)</td>
</tr>
<tr>
<td>3' AAGGGAGGTCGAGGATTCGCCACCG (nt 577–600)</td>
<td>5' ATGGGCCCCCTGGAGGTCCAAGGACC (nt 645–678)</td>
</tr>
<tr>
<td>Fragment 2</td>
<td></td>
</tr>
<tr>
<td>5' TGGTCGCCCCCGGAGGGGGGCGGA (nt 443–467)</td>
<td>5' TGGAGCTTATCCCGGGGATCG (nt 469–491)</td>
</tr>
<tr>
<td>3' TGTCACTCCTCGAGTGATCCCA (nt 958–982)</td>
<td>5' CAGGGTTTCACCTACGGTTCCGCTC (nt 922–948)</td>
</tr>
<tr>
<td>Fragment 3</td>
<td></td>
</tr>
<tr>
<td>5' GATGCCCCAGTGCGACCGGGAGAGG (nt 855–879)</td>
<td>5' GGGAGATGTCGAGGGCCGGAAAGAG (nt 882–906)</td>
</tr>
<tr>
<td>3' AAAAGGGAGGAGCAGGGGGGAGGG (nt 1394–1421)</td>
<td>5' GGCGAGGAGCCGAGCGAGGATCAG (nt 1364–1388)</td>
</tr>
<tr>
<td>Fragment 4</td>
<td></td>
</tr>
<tr>
<td>5' GAGATCCCTCCTCCTCCTTTGCTGGT (nt 1292–1316)</td>
<td>5' GTGAGGCGCTTCCAGGGCCGGA (nt 1328–1351)</td>
</tr>
<tr>
<td>3' TCTGTGAATGAAATTCCGGGAGTCTC (nt 146–173)</td>
<td>5' CAGGGTTCCGAGGAAATCTCCTC (nt 125–145)</td>
</tr>
</tbody>
</table>
HDV genotype Ib variant

Fig. 1. For legend see page 3278.
analysis of the partial HDV genome encompassing the HDAg region, including seven additional partial sequences determined in this study, was also performed.

**Statistical analysis.** Categorical data were compared by chi-square or Fisher’s exact test. Distributions of continuous variables were analysed by the Mann–Whitney U-test or Student’s t-test using Statview 5.0 software (Abacus Concepts). All tests of significance were two-tailed and \( P \) values of less than 0.05 were considered as statistically significant.

## RESULTS

### Homology and phylogenetic tree analysis of the HDV genome

Fig. 1 shows the neighbour-joining phylogenetic tree based on the complete sequences of 55 HDV isolates, including the 33 newly sequenced HDV isolated from the present study and another 22 available full-length genome sequences. Of the 33 new HDV isolates, two were located in the genotype I cluster and one in the genotype IIa. The two patients with genotype I were from Nagasaki and the patient with genotype IIa was from Miyako Islands, Okinawa. The other 30 strains clustered with the reported complete sequences of genotype IIb TWD62 or L215 (Wu et al., 1998; Ma et al., 2003). Six strains clustered with the prototype genotype IIb isolate TWD62, originally reported from Taiwan, whereas the other 24 isolates formed a unique cluster with the IIb strain L215, a recently reported full genome sequence from the Miyako Islands (Ma et al., 2003). The two clusters were divided distinctly with high bootstrap values of 100 %. Five of the six strains clustered with the prototype genotype IIb were from the Miyako Islands, while the other patient with prototype genotype IIb was from Tokyo. In contrast, the strains from the cluster including L215 were all from the Miyako Islands. These results indicate the existence of at least three lineages of HDV variant, Ila, prototype IIb and Maiyako Islands-specific variant genotype IIb-Miyako (IIb-M) in the Miyako Islands. A maximum-likelihood tree constructed from the same sequences also showed distinct clusters of prototype IIb and IIb-M with high bootstrap values (data not shown).

The typical genomic sequences of genotype IIb-M with representative isolates of other HDV genotypes (I, Ila, IIb, III) are shown in Fig. 2. The length of the complete HDV sequence of genotype IIb-M was 1676 nt and the overall identities in the HDV genomic sequences between genotype IIb-M and genotype I (Wang et al., 1986), Ila (Imazeki et al., 1991), IIb (Wu et al., 1998), IIb-L215 (Ma et al., 2003) and III (Casey et al., 1993) isolates were 72-7, 78-1, 87-3, 93-8 and 64-6 %, respectively, showing that IIb-M is most closely related to IIb-L215, the recently reported HDV isolate from the Miyako Islands.

The lengths of the complete HDV sequences of prototype IIb and of IIb-M genotypes determined in this study were 1677–1679 nt and 1675–1685 nt, respectively. The nucleotide identities among isolates within prototype IIb and within genotype IIb-M were 94–97 % similar, whereas those between prototype IIb and IIb-M had similarities of 88–90 %. There were marked variations in the degree of genetic divergence among different regions of the HDV genome when comparing IIb-M (JA-M1) and the other genotypes (Modahl & Lai, 2000). As shown in Table 2, the greatest divergence was in the hypervariable region (54-7–92-5 %), followed by the HDAg coding sequence (70-7–94-4 %), and the least divergence was in the autocleavage region (74-8–95-6 %).

### Clinical characteristics of patients with HDV genotype IIb-M

A phylogenetic tree analysis of the partial genome sequences encompassing the HDAg coding region, including additional sequences isolated from seven patients from the Miyako Islands, showed similar clustering (data not shown). Five of them clustered with genotype IIb-M and the remaining two strains clustered with prototype IIb. Finally, the clinical pictures between eight patients infected with prototype IIb and 29 patients with IIb-M were compared. As shown in Table 3, age, sex, HBV status [all patients were HBeAg negative and anti-HBe positive with low \( <1-0 \times 10^5 \) copies (ml serum)\(^{-1}\) HBV DNA levels] did not differ between these two groups. However, the severity of liver disease was significantly higher in patients with genotype IIb-M than in those with prototype IIb. In prototype IIb patients, three were ASC, five were CH and none were LC. In contrast, there were 16 CH and 13 LC in genotype IIb-M. Thus, patients with HDV IIb-M showed significantly greater disease progression compared with patients with HDV prototype IIb (\( P=0-0009 \) by the Kruskal Wallis test).

### Genetic features of HDV genotype IIb-M

The secondary structure of the antigenic sequence corresponding to the 3’ end of the small HDAg gene containing the RNA editing site (Casey et al., 1992) showed that IIb-M has a particular structure located at nt 1012, the amber/tryptophan site (Fig. 3). All known HDV genotypes I

---

**Fig. 1.** Phylogenetic tree analysis of HDV isolates. Sources of isolate sequences are as follows: TWD62, AF018077; Taiwan-3, U19598; Taiwan-1, M92448; Yakut-26, AJ309879; Yakut-62, AJ309880; Japan-1, X60193; Nauru, M58629; Lebanon-1, M84917; Somalia, U81988; China, X77627; US-1 (M28267; US-2, L22066; France, D01075; Italy-1, X04451; Italy-2, X85253; Canada, AF098261; Central African Republic, AJ000558; Peru-1, L22063; VnzD8349, AB037948; VnzD8375, AB037947; VnzD8624, AB037949. JA-M1 to JA-M37 (from Miyako), JA-N1 and JA-N2 (from Nagasaki) and JA-T (from Tokyo) were sequenced in this study. These sequences have been deposited in the GenBank database (AF309420 and AB118818–AB118849).
HDV genotype IIb variant

Fig. 2. For legend see page 3282.
Fig. 2. (cont.) For legend see page 3282.
Fig. 2. (cont.) For legend see page 3282.
and II, including the prototype genotype IIb, have a 2–4 bp structure on both sides of the RNA editing site, located on the opposite strand at around nt 580, and this structure is required for efficient RNA editing and HDV replication (Nakano et al., 2001; Ma et al., 2003; Ivaniushina et al., 2001). The RNA editing site of IIb-M showed a unique structure, where this base-paired structure was disrupted one nucleotide upstream adjacent to the RNA editing site. Furthermore, these structures around the RNA editing site and the opposite site were well conserved among genotype IIb and among genotype IIb-M, despite being different between genotypes IIb and IIb-M.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>IIb-M (L215)</th>
<th>IIb (TWD62)</th>
<th>Ila (Japan-1)</th>
<th>I (Italy-1)</th>
<th>III (Peru-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sequence (%)</td>
<td>93:8</td>
<td>87:3</td>
<td>78:1</td>
<td>72:7</td>
<td>64:6</td>
</tr>
<tr>
<td>Autocleavage region (nt 658–956)</td>
<td>95:6</td>
<td>93:0</td>
<td>92:9</td>
<td>89:2</td>
<td>74:8</td>
</tr>
<tr>
<td>Delta antigen (nt 957–1597)</td>
<td>94:4</td>
<td>90:1</td>
<td>81:5</td>
<td>77:9</td>
<td>70:7</td>
</tr>
<tr>
<td>Hypervariable region (nt 1598–657)</td>
<td>92:5</td>
<td>82:5</td>
<td>69:0</td>
<td>61:1</td>
<td>54:7</td>
</tr>
</tbody>
</table>

**Table 3. Clinical and virological features according to HDV genotype**

HBeAg was negative in all the patients.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>I (n = 2)</th>
<th>Ila (n = 1)</th>
<th>IIb (n = 8)</th>
<th>IIb-Miyako (n = 29)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (male : female)</td>
<td>0 : 2</td>
<td>1 : 0</td>
<td>4 : 4</td>
<td>18 : 11</td>
<td>NS</td>
</tr>
<tr>
<td>Age (years)</td>
<td>34 ± 16</td>
<td>39</td>
<td>59 ± 14</td>
<td>61 ± 12</td>
<td>NS</td>
</tr>
<tr>
<td>ALT (IU l⁻¹)</td>
<td>57 ± 7</td>
<td>63</td>
<td>52 ± 27</td>
<td>94 ± 182</td>
<td>NS</td>
</tr>
<tr>
<td>HBV genotype</td>
<td>A 1/B 1</td>
<td>B 1</td>
<td>B 8</td>
<td>B 29</td>
<td>NS</td>
</tr>
<tr>
<td>HBV DNA level*</td>
<td>3-7 (&lt;3-7)</td>
<td>&lt;3-7†</td>
<td>&lt;3-7 (&lt;3-7-4-1)†</td>
<td>&lt;3-7 (&lt;3-7-5-5)†</td>
<td>NS</td>
</tr>
<tr>
<td>Stage (ASC : CH : LC)</td>
<td>0 : 1 : 1</td>
<td>0 : 1 : 0</td>
<td>3 : 5 : 0</td>
<td>0 : 16 : 13</td>
<td>0 : 0009‡</td>
</tr>
<tr>
<td>HCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>NS</td>
</tr>
<tr>
<td>Distribution</td>
<td>Nagasaki 2</td>
<td>Miyako 1</td>
<td>Miyako 7</td>
<td>Miyako 29</td>
<td>NS</td>
</tr>
</tbody>
</table>

*log10 HBV DNA (copies ml⁻¹).
†Median range.
‡Between IIb and IIb-Miyako.

**Deduced HDAg sequence of HDV genotype IIb in Okinawa**

Fig. 4 shows the predicted HDAg amino acid sequence of the representative isolate of HDV genotype IIb-M aligned with HDV isolates of other genotypes. The identities in the amino acid sequences of the large HDAg between IIb-M and I (Wang et al., 1986), IIa (Imazeki et al., 1991), IIb (Wu et al., 1998), IIb-M (L215) (Ma et al., 2003) and III (Casey et al., 1993) were 70, 79, 82 and 61 %, respectively (Table 4). There was considerable variation among genotypes within several domains: the RNA-binding domain was the most...
conserved region, followed by the nuclear localization signal (NLS) region, the coiled-coil sequence containing the leucine zipper motif and the N terminus, and the C-terminal packaging sequences of 19 amino acids (Modahl & Lai, 2000). The cysteine residue four amino acids from the C terminus, the site of prenylation required for binding with HBsAg (Glenn et al., 1992), was well conserved in every isolate. Comparing IIb and IIb-M, the NLS and RNA binding domains were well conserved, whereas the coiled-coil domain and packaging sequences showed significant differences. In the coiled-coil domain, the substituted amino acid residues had similar properties (D → E, T → N and I → L) and the leucine zipper motif was completely conserved (Chen et al., 1992; Wang & Lemon, 1993), while the packaging sequence of most of the IIb-M isolates showed the characteristic four amino acid substitutions in the proline.

Fig. 3. RNA editing site of HDV genotype IIb and IIb-M. (a) Nucleotide sequences of the RNA editing site of HDV genotypes IIb and IIb-M, which is formed between anti-genome RNA surrounding the edited A residue (nt 1012) and nucleotide sequences of the opposite site (nt 580) of the unbranched rod structure of HDV. Sources of isolate sequences are as follows: L215, AB088679; TWD62, AF18077. JA-M1 to JA-M36 (from Miyako) and JA-T (from Tokyo) were sequenced in this study.

(b) The base-paired structure formed by genotype IIb, IIb-M, I and III for RNA editing.
Fig. 4. HDAg amino acid alignment of HDV isolates. The amino acids are numbered according to Wang et al. (1986). Dots indicate conserved amino acids and dashes indicate missing amino acids. Sources of isolate sequences are as follows: L215, AB088679; TWD62, AF18077; Japan-1, X60193; Yakut-26, AJ309879; China, X77627; Italy-1, X04451; Nauru, M58629; Peru-1, L22063. JA-M1 and JA-M37 (from Miyako), JA-N1 (from Nagasaki) and JA-T (from Tokyo) were sequenced in this study.
Table 4. Amino acid identities of HDAg between IIb-Miyako and other genotypes

Numbers given are the amino acid identities of HDAg (%) between the isolate JA-M1 (AF309420) and each isolate listed below. Sources of the isolates: L215, AB088679; TWD62, AF18077; Japan-1, X60193; Italy-1, X04451; Peru-1, L22063.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Complete sequence (%)</th>
<th>Coiled-coil domain (aa 31–52)</th>
<th>Nuclear localization domain (aa 68–88)</th>
<th>RNA-binding domain (aa 97–146)</th>
<th>Packaging signal domain (aa 195–214)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIb-M (L215)</td>
<td>92</td>
<td>86</td>
<td>100</td>
<td>88</td>
<td>84</td>
</tr>
<tr>
<td>IIb (TWD62)</td>
<td>87</td>
<td>77</td>
<td>100</td>
<td>96</td>
<td>79</td>
</tr>
<tr>
<td>IIa (Japan-1)</td>
<td>79</td>
<td>68</td>
<td>86</td>
<td>84</td>
<td>68</td>
</tr>
<tr>
<td>I (Italy-1)</td>
<td>70</td>
<td>59</td>
<td>95</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td>III (Peru-1)</td>
<td>61</td>
<td>64</td>
<td>62</td>
<td>78</td>
<td>64</td>
</tr>
</tbody>
</table>

Fig. 5. C-terminal end of HDAg of HDV genotypes IIb and IIb-M. The C terminus of the HDAg from all the IIb (n=8) and IIb-M (n=29) genotypes were aligned. Dots indicate the same amino acids as the consensus sequence of genotype IIb. The stage (ASC, CH or LC) is described at the right side of the sequences. An asterisk indicates a proline residue in the packaging signal. Sources of isolates are as follows: L215, AB088679; TWD62, AF18077. JA-M1 to JA-M36 (from Miyako) and JA-T (from Tokyo) were sequenced in this study.
stretch that regulates extranuclear export of HDAg (Lee et al., 2001) (Fig. 5).

**DISCUSSION**

In the present study, we initially sequenced the full-length HDV genome from 33 patients in Japan with chronic HDV infection and demonstrated a new HDV genotype IIb variant in the Miyako Islands, Okinawa, the Japanese islands nearest to Taiwan. Patients infected with genotype IIb-M showed greater progression to CH or liver cirrhosis than genotype IIb-infected patients with similar clinical backgrounds. HDV genotype IIb-M has specific genetic structures in the RNA editing site and the packaging signal sequence of HDAg, which could potentially influence the efficiency of HDV replication (Casey, 2002; Hsu et al., 2002; Lee et al., 1995, 2001; Wu et al., 1995b; Yang et al., 1995). The observed correlation between HDV genetic structure and clinical characteristics suggests a critical role for variations in the RNA editing site and packaging signal of the HDAg gene in determining the diversity of clinical outcome, even among patients infected with the same genotype of HDV.

We identified the new HDV genotype IIb variant by phylogenetic analysis of the complete genomes of 33 HDV isolates. Among them, 30 isolates, mostly from the Miyako Islands, were classified as genotype IIb (Wu et al., 1998) or its variant, IIb-M. In previous studies including our own, HDV genotypes in the Miyako Islands have been considered as genotype IIb (Sakugawa et al., 1999; Ma et al., 2003; Arakawa et al., 2000). However, the present detailed phylogenetic analysis using the full genome successfully identified a cluster distinct from the prototype IIb cluster. In fact, the nucleotide homologies between genotype IIb and IIb-M and among genotype IIb-M were clearly different, i.e. 88–90% and 94–97%, respectively. HDV genotype II is divided into two types in Taiwan (i.e. IIa and IIb), with 77% nucleotide homology between the complete sequences of genotype IIa and IIb (Wu et al., 1998). Although the criteria for defining identical genotype by homology analysis were not determined, the difference between IIb and IIb-M seems to be less than that between IIa and IIb, as shown by phylogenetic tree analysis. In fact, a IIa variant was recently reported in Yakutia, Siberia, Russia, also causes a severe hepatitis comparable with genotype I in this cohort (Ivaniushina et al., 2001). These findings strongly suggest that the genetic structure of HDV can profoundly influence the pathogenesis of liver injury in HDV infection. However, the genetic structure responsible for such clinical features could not be readily determined because the genetic differences between the different genotypes are too diverse, as seen in Fig. 2. In contrast, despite the different clinical pictures between IIb and IIb-M, the genetic differences are small enough to enable the definition of the genetic features of HDV pathogenesis and replication in vivo.

One of the most important findings in the present study is that the clinical pictures differ between genotype IIb and IIb-M. Our previous studies demonstrated that HDV genotype II is predominant in this area and that these patients show heterogeneous clinical pictures ranging from ASC to HCC (Sakugawa et al., 1999; Nakasone et al., 1998); however, the reason for this diversity could not be explained based on the known clinical and virological factors of HBV. In the present study, all of the patients with chronic HDV genotype IIb infection were ASC or CH and none were at the LC or HCC stage. In contrast, 55% and 45% of patients with genotype IIb-M were in the CH and LC stages, respectively, and none of them was ASC. These findings indicate that patients with genotype IIb-M are more likely to progress to LC and HCC than those with genotype IIb and that differences in HDV genotype could cause the different clinical pictures observed in this population.

The main cause of the difference in liver disease between patients with IIb and IIb-M seems to be the diversity of HDV itself. Although the severity of liver disease in hepatitis D can be influenced by a variety of host factors including genetic backgrounds as well as HBV status, no apparent differences were found between patients with genotype IIb and IIb-M. In particular, in most patients, serum HBV DNA levels were below $10^5$ copies ml$^{-1}$ with negative HBeAg, which were too low to cause HBV-related liver injury (Sakugawa et al., 2001; Lok et al., 2001). Similarly, the HBV genotype, which is also known to cause diversity of liver disease (Kao et al., 2000; Orito et al., 2001), was genotype B in all of the patients from the Miyako Islands. Differences in HDV genotype are known to affect the pathogenesis and diverse clinical pictures of HDV infection (Casey et al., 1993; Wu et al., 1995a; Ivaniushina et al., 2001). Genotype III, exclusively found in the northern part of South America, is associated with fulminating hepatitis (Casey et al., 1993). On the other hand, genotype II in Taiwan is generally associated with a more favourable outcome than genotype I, which causes liver disease with diverse clinical presentation from asymptomatic carrier to rapidly progressive CH (Wu et al., 1995a). A IIa variant recently reported in Yakutia, Siberia, Russia, also causes a severe hepatitis comparable with genotype I in this cohort (Ivaniushina et al., 2001). These findings strongly suggest that the genetic structure of HDV can profoundly influence the pathogenesis of liver injury in HDV infection. However, the genetic structure responsible for such clinical features could not be readily determined because the genetic differences between the different genotypes are too diverse, as seen in Fig. 2. In contrast, despite the different clinical pictures between IIb and IIb-M, the genetic differences are small enough to enable the definition of the genetic features of HDV pathogenesis and replication in vivo.

By comparative analysis between the genotype IIb and IIb-M genomes, the highest difference was found in the
hypervariable region (nt 1598–657) and moderately high in HDAg (nt 957–1597), whereas the autocatalytic regions encoding ribozyme activity were well conserved (Wu & Lai, 1989). The hypervariable region was markedly variable even within the same genotype, supporting the notion that this region cannot confer any relevant biological function aside from the formation of the rod structure of HDV RNA required for RNA synthesis by RNA polymerase II (Modahl & Lai, 2000). On the other hand, the requirement for strict secondary or tertiary structure of the autocatalytic domain seems to be so crucial for full activity of the ribozyme needed for the rolling-circle mechanism of HDV replication that divergence of this region could not exist among isolates. Therefore, HDV genetic regions other than the hypervariable region or the autocatalytic domain, i.e. the HDAg coding region, confer the clinical difference between IIb and IIb-M. In the HDAg coding region, we found that the most prominent differences are in the RNA editing site and the packaging signal in the C terminus of the large HDAg (Modahl & Lai, 2000). Although the coiled-coil domain (Wang & Lemon, 1993) also showed modest differences, the leucine zipper motif (Chen et al., 1992) was preserved, and the nuclear localizing signal (Xia et al., 1992) and RNA binding domain (Lin et al., 1990) were identical in IIb and IIb-M, indicating that these regions are not responsible for liver damage.

In genotype IIb-M, there was particular disruption of the base-pairing structure two bases upstream of the editing site, resulting in a characteristic structure in this region distinct from genotype IIb and IIa (Fig. 3). There is a possibility that the unique structure of the RNA editing site of genotype IIb-M may affect the observed difference in pathogenesis between genotype IIb and IIb-M. RNA editing is a pivotal event during the HDV replication cycle (Casey et al., 1992), where initially in HDV infection, small HDAg transactivates HDV RNA synthesis by RNA polymerase II (Wu et al., 1995b). Large HDAg, which has 19 additional amino acids (the packaging signal sequence) at the C terminus of small HDAg, is produced in the late stage of infection by RNA editing of the amber stop codon (UAG) to a tryptophan codon (UGG) in the small HDAg gene by the host adenosine deaminase (Modahl & Lai, 2000). Large HDAg suppresses HDV RNA replication and promotes virion assembly by extranuclear export of the HDAg–RNA complex and binding to HBsAg. The regulatory mechanism of this RNA editing is not fully understood, but the secondary structure of the antigenomic region corresponding to the 3′ end of the small HDAg gene influences the editing efficiency (Casey et al., 1992; Wu et al., 1995b; Casey, 2002). A recent in vitro mutational study clearly demonstrated that the base-pairing structure surrounding the RNA editing site profoundly influences RNA editing efficiency (Hsu et al., 2002). In genotype I, the base pairing surrounding this site is particularly strong (four base pairs on each side), whereas a weaker secondary structure is found within genotype II that is associated with milder liver disease. In addition, the distinct structure of genotype III is thought to be involved in fulminant hepatitis (Casey, 2002). Collectively, the specific differences in the base-paired structure of the RNA editing site might explain to some extent the difference in virulence among HDV genotypes. Therefore, although in vitro confirmation is necessary, it appears that the loose structure around the RNA editing site found in genotype IIb-M might influence the editing efficiency in comparison with genotype IIb, leading to the observed clinical differences.

In addition to the difference in the RNA editing site, there are four characteristic amino acid differences (codons 198, 200, 201 and 203) in the packaging signal sequence of the large HDAg between genotype IIb and IIb-M (Fig. 5). This region is almost completely conserved among IIb-M isolates. As mentioned above, addition of this packaging signal reverses the property of HDAg (Modahl & Lai, 2000; Chang et al., 1993). The exact molecular mechanism of this phenomenon is not completely understood, but, as shown in Fig. 4, a sequence of 19 amino acids was highly genotype specific. In vitro analysis demonstrated that swapping the packaging signal sequence of genotype IIa with that of genotype I HDAg decreases the virus replication of genotype I, while the replication of genotype II was intensified, indicating that this region directly regulates HDV RNA replication (Hsu et al., 2002). Thus, the structural characteristics of this region in IIb-M can profoundly influence virus replication. In particular, two of the four amino acid differences found in IIb-M were located in the proline residues, which are implicated in the assembly process by extranuclear export of the HDAg–RNA complex. In fact, in a recent study with cultured cells, mutation of the proline residue in this region attenuated the extranuclear export of large HDAg (Lee et al., 2001). However, these data did not directly prove that the C-terminal domain structure of HDAg influences the pathogenesis. In the future, in vitro mutational studies should be performed to verify the hypothesis that differences in the packaging signal sequence in genotype IIb-M can modulate HDV replication and lead to progressive disease.

In conclusion, we have identified a new genetic subclass of HDV genotype IIb in the Miyako Islands, Okinawa, Japan. This HDV variant is associated with more aggressive liver disease and has specific genetic changes in the C-terminal packaging signal of large HDAg as well as the RNA editing sequence. These findings should prompt further investigation into the relationship between HDV genetic structures and their function and pathogenesis. This study provides valuable information for molecular epidemiology and diagnosis, contributes to a better understanding of HDV biology and offers the potential for new therapies for HDV: a disease for which no effective therapy has yet been established.

**ACKNOWLEDGEMENTS**

This work was supported in part by a Grant-in-aid (08457164) from the Ministry of Education, Sciences, Culture and Sports of Japan.
REFERENCES


