Characterization and phylogenetic analysis of a novel hepatitis D virus strain discovered by restriction fragment length polymorphism analysis

Jaw-Ching Wu,1 Tzen-Yuh Chiang2 and I-Jane Sheen1

1 Division of Gastroenterology, Department of Medicine, Veterans General Hospital, National Yang-Ming University School of Medicine, Taipei 112, Taiwan, Republic of China
2 Department of Biology, Cheng-Kung University, Tainan, Taiwan, Republic of China

The hepatitis D virus (HDV) genotypes in 46 HDV-infected patients and 12 prostitutes were screened with Xhol restriction fragment length polymorphism (RFLP) analysis of reverse transcription PCR products of viral genomes and verified by phylogenetic analysis. The amplificates of three (6.5%) patients and two (17%) prostitutes showed a novel RFLP pattern different from those of the three known genotypes. Complete HDV genomic sequence identities between isolates with a novel RFLP and the HDV genotypes I, II and III were 72.3, 77.2 and 63.0%, respectively. Importantly, divergence was mostly seen in various regions related to replication or packaging. The novel isolates formed a monophyletic group (P < 0.05) and were most closely related to genotype II.

Introduction

Hepatitis D virus (HDV), a defective virus, requires the supply of hepatitis B surface antigen (HBsAg) envelope by hepatitis B virus (HBV) for its assembly and transmission (Bonino et al., 1986; Wu et al., 1991; Rizzetto et al., 1980). Delta antigen is the only protein encoded by HDV. There are two forms of delta antigen; the large delta antigen has an extra 19 amino acids at its C terminus compared to the small form (Bergmann et al., 1986; Weiner et al., 1988). The small delta antigen is essential for HDV replication, whereas the large form is responsible for virus assembly and trans-dominant inhibition of virus replication (Kuo et al., 1989; Chang et al., 1991; Chen et al., 1992).

The disease spectrum of HDV infection ranges from fulminant hepatitis and rapid cirrhotic progression to a subclinical course (Govindarajan et al., 1984, 1986; Wu et al., 1994a, 1995a). HDV is currently classified into three genotypes based on a comparison of a segment of sequence within the delta antigen-coding region in the antigenomic strand of HDV RNA (Casey et al., 1993). It has been recently reported that HDV genotype I is more often associated with grave outcomes than genotype II (Wu et al., 1995b). HDV genotype III is frequently associated with fulminant hepatitis (Casey et al., 1993, 1996). However, the clinical manifestations vary greatly even in infections of the same genotype (Wu et al., 1995b; Niro et al., 1997). HDV shows 11−19% heterogeneity in nucleotide sequences of the same genotype (Casey et al., 1993; Chao et al., 1991; Wu et al., 1995c). There is 27−34% divergence in nucleotide sequences and 30% divergence in the amino acid sequences among different genotypes (Casey et al., 1993; Chao et al., 1991; Wu et al., 1995c). In vitro mutagenesis of various regions of the delta antigen-coding sequence influences the replication or the packaging of HDV (Chen et al., 1992; Glenn et al., 1992; Lee et al., 1993, 1994; Lazinski & Taylor, 1993; Lai, 1995); thus, the naturally occurring HDV genomic variations in this region may also influence virus behaviour and subsequently influence the clinical course.

HDV genotype I is widespread in Italy, the United States, Taiwan, Nauru, France, Lebanon and China (Niro et al., 1997; Chao et al., 1990, 1991; Wang et al., 1986; Makino et al., 1987; Kuo et al., 1988; Saldanha et al., 1990; Lee et al., 1992; Zhou et al., 1994). Genotype II has been isolated only in Japan and Taiwan (Wu et al., 1995b, 1995c; Imazeki et al., 1990; Lee et al., 1996). Genotype III has been isolated only in northern South America (Casey et al., 1993, 1996). It is not known if there are any further HDV genotypes. It is too tedious to search for novel HDV strains by cloning or direct sequencing from a large number of samples. In this study, we used a genotyping method based on restriction fragment length polymorphism (RFLP) analysis (Wu et al., 1995b) and discovered a novel RFLP pattern different from those of the three known genotypes. Complete HDV genomic sequence identities between isolates with a novel RFLP and the HDV genotypes I, II and III were 72.3, 77.2 and 63.0%, respectively. Importantly, divergence was mostly seen in various regions related to replication or packaging. The novel isolates formed a monophyletic group (P < 0.05) and were most closely related to genotype II.
pattern. The HDV isolates with a novel RFLP pattern were cloned, characterized and analysed phylogenetically. Various regions of the HDV genome related to replication or packaging were compared in the novel isolate and the three known genotypes.

**Methods**

| HDV-infected patients and prostitutes. A total of 46 HBV carriers with acute HDV superinfection (eight cases with fulminant hepatitis and 38 cases with acute non-fulminant hepatitis), who had been tested positive for serum HDV RNA by RT–PCR, were included for further analysis of their HDV genotype. They were all positive for serum HBsAg and antibody to HDV antigen and negative for IgM antibody to hepatitis B core antigen (Abbott Laboratories). The diagnostic criteria for acute HDV superinfection were described in our previous reports (Wu et al., 1994a, 1995a). According to our previous report (Wu et al., 1990), sexual contact with prostitutes is the most common transmission route of HDV infection in Taiwan. Therefore, 12 serum HDV RNA-positive prostitutes, described in our previous study (Wu et al., 1994b), were included.

| RT–PCR and RFLP analysis. RT–PCR, using primers #120 (homologous to a sequence from nt 889–912), #214 and #214′ (complementary to a sequence from nt 1334–1313, according to the corresponding delta antigen-coding region in the antigenomic strand of HDV RNA), was performed as reported previously, with some modification according to the sequence of genotypes I or II (Chao et al., 1991; Wu et al., 1995c). The PCR products contained an HDV genomic region from nt 911–1260, a region that has been proposed for the classification of HDV genotypes by Casey et al. (1993). The phylogenetic tree based on this region is similar to those based on the whole HDV genome or the hypervariable region (Casey et al., 1993). The PCR products cleaved to a novel RFLP pattern were then cloned and sequenced. The RFLP patterns had perfect agreement with genotyping by phylogenetic analysis based on sequence comparison and phylogenetic analysis. For the cloning of the whole HDV genome with the novel RFLP pattern, two sets of primers, antisense 

| Phylogenetic analysis. HDV sequences were multiple aligned using the CLUSTAL V program of Higgins et al. (1992). Phylogenetic analysis using parsimony (PAUP) (version 3.1.1; Swofford, 1993), the neighbour-joining method [molecular evolutionary genetics analysis (MEGA) program, version 1.01; Kumar et al., 1993] and PHYLIP (version 3.5) were used for phylogenetic analysis of sequencing data (Swofford et al., 1996). Neighbour-joining analyses were conducted by calculating Kimura’s two-parameter distance (Kimura, 1980). Sequences of South American isolates, i.e. Peru1, Peru2 and Colombia (Casey et al., 1993), were chosen as outgroups based on a previous study and biogeographical evidence. The confidence of the clades, i.e. the monophyly, was tested by bootstrapping with 1000 replicates of heuristic searches. The nodes with bootstrap values greater than 70% are significantly supported with > 95% confidence (robustness) (Hillis & Bull, 1993).

**Results**

Discovery of a novel RFLP pattern of PCR products of HDV genomes in patients

HDV genomes from sera of 46 patients were reverse transcribed and amplified by PCR. The size of the PCR products of the novel HDV isolate have a single-base deletion at 5′ end of the amplified sequence; 3. PCR products of the novel HDV isolate have a single Xhol cutting site and are cleaved into fragments of 83 and 362 bp (†, the novel HDV isolate has a single-base deletion at 5′ end).
Cloning and sequencing of HDV with a novel RFLP pattern

The PCR products derived from HDV genomes from the three patients with acute hepatitis and the two asymptomatic prostitutes that showed a novel RFLP pattern were cloned and sequenced. These five subjects were not related to each other. The primers #94 and #1622 were used to amplify a 775 base cDNA fragment that includes the complete delta antigen-coding region. Based on the sequence obtained from clones containing this fragment, primers #945 and #1622 were synthesized and used to amplify a fragment of 1000 bases at the other end of the HDV genome. There are 42 base and 57 base overlaps at the ends of these fragments. At least five clones were obtained from separate PCRs for each fragment to reduce errors possibly introduced by PCR. The novel HDV RNA, designated TW (IIb) (GenBank accession no. AF018077), has 1676 nt in total. The heterogeneity in the nucleotide sequences of clones from a single subject was 0.29–1.1%, while the novel HDV clones of the five subjects showed >97% identity with each other.

The aligned HDV genomic sequences of one of the novel isolates and representative isolates of three HDV genotypes are shown in Fig. 2. The overall identities in the HDV genomic sequences between the novel isolate and the genotype I, II and III isolates were 72.3, 77.2 and 63.0%, respectively. There are marked variations in divergence in different regions of the HDV sequence among the novel isolates and the different genotypes. In the region from nt 911–1260, the novel isolates had identities of 76.0–78.3% with genotype I isolates, 82.3–85.4% with genotype II isolates, and 67.4% with the genotype III isolate. As shown in Table 1, the greatest divergence was in the hypervariable region (30.8–45.7%), followed by the delta antigen-coding sequence (20.2–30.7%) and the least divergence was in the autocleavage region (8.7–28.8%). Within the delta antigen-coding sequence, the greatest divergence (21–61%) was from nt 960–1016, which corresponded to the carboxyl end of the open reading frame for the large delta antigen on the antigenomic strand of RNA (Fig. 2).

The predicted amino acid sequences of the delta antigen of the novel isolate and the three genotypes are shown in Fig. 3. The identities in the amino acid sequences of the large delta antigen between the novel isolate and the three genotypes were 69.9, 78.2 and 62.0%, respectively. There are great variations in identity in various domains among the novel isolate and the three genotypes: the RNA-binding domain was the most conserved region, followed by the nuclear localization signal region, the coil–coil sequence, the N terminus and the C-terminal packaging sequence (Table 2). The cysteine residue near the C terminus is well-conserved in every isolate. Three short segments of amino acids (aa 100–114, aa 124–132 and aa 159–168) are completely conserved among the novel isolate and all genotypes. Of note, the novel isolate showed a higher identity to genotype II at both ends of the delta antigen, but a

<table>
<thead>
<tr>
<th>Isolate (genotype)</th>
<th>Italy (I)</th>
<th>Japan-1 (IIa)</th>
<th>Peru (III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sequence</td>
<td>72.3</td>
<td>77.2</td>
<td>63.0</td>
</tr>
<tr>
<td>Autocleavage region (nt 659–959)</td>
<td>87.7</td>
<td>91.3</td>
<td>71.2</td>
</tr>
<tr>
<td>Delta antigen coding region (nt 960–1601)</td>
<td>76.1</td>
<td>79.8</td>
<td>69.3</td>
</tr>
<tr>
<td>Hypervariable region (nt 1602–658)</td>
<td>62.7</td>
<td>69.2</td>
<td>54.3</td>
</tr>
</tbody>
</table>

Table 1. Comparison of various regions of whole genomic sequences between a novel isolate (TWD62) and isolates of three HDV genotypes

Numbers given are the nucleic acid identities (%) between the isolate TWD62 and each isolate listed. Sources of the isolates: Italy, Wang et al. (1986); Japan-1, Imazeki et al. (1990); Peru, Casey et al. (1993). For various regions of HDV, see review by Lai (1995).
higher identity with genotype I in the middle part of the delta antigen (the nuclear localization signal and the RNA-binding regions).

Phylogenetic analysis of HDV isolates

Phylogenetic analysis was conducted on ten previously published HDV genotype I isolates, two published genotype II...
Fig. 2. Alignments of whole genomic sequences of representative HDV isolates of three genotypes and a novel strain. The numbering of nucleotides is according to Makino et al. (1987). The corresponding initiation and termination codons of the open reading frame for the small and the large delta antigens in the antigenomic sense of HDV RNA are indicated by transverse arrows and vertical bars associated with the block letters HDAg, S and L. The nucleotide sequence of the autocatalytic region is underlined according to Chao et al. (1991). The genomic autocatalytic cleavage site (688/689) is indicated by ‘$’’. The site (903/904) corresponding to the cleavage site of the antigenomic RNA is indicated by an arrow head. Dots indicate conserved nucleotides and dashes indicate missing nucleotides. Figures in parentheses indicate genotype. Sources of isolates are indicated: TWD62, our laboratory (GenBank accession no. AF018077); J, Japan-1 (Imazeki et al., 1990); I, Italy (Wang et al., 1986); and P, Peru1 (Casey et al., 1993).
Table 2. Comparison of various domains of the amino acid sequences of the delta antigens between a novel isolate (TWD62) and isolates of three HDV genotypes

Numbers given are the amino acid identities (%) between the delta antigen of isolate TWD62 and each isolate listed. Sources of the isolates: Italy, Wang et al. (1986); Japan-1, Imazeki et al. (1990); Peru, Casey et al. (1993). For various domains of the hepatitis delta antigen, see review by Lai (1995).

<table>
<thead>
<tr>
<th>Isolate (genotype)</th>
<th>Italy (I)</th>
<th>Japan-1 (IIa)</th>
<th>Peru (III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sequence</td>
<td>69.9</td>
<td>78.2</td>
<td>62.0</td>
</tr>
<tr>
<td>N terminus (aa 1–30)</td>
<td>50.0</td>
<td>63.3</td>
<td>36.7</td>
</tr>
<tr>
<td>Coi-coil sequence (aa 31–52)</td>
<td>63.6</td>
<td>77.3</td>
<td>68.2</td>
</tr>
<tr>
<td>Nuclear localization domain (aa 68–88)</td>
<td>100</td>
<td>90.4</td>
<td>66.6</td>
</tr>
<tr>
<td>RNA-binding domain (aa 97–146)</td>
<td>94.0</td>
<td>88.0</td>
<td>86.0</td>
</tr>
<tr>
<td>Packaging signal (aa 195–214)</td>
<td>35.0</td>
<td>85.0</td>
<td>40.0</td>
</tr>
</tbody>
</table>

Fig. 3. The predicted amino acid sequences of the novel HDV isolate and the representative isolates of the three known genotypes. Dots indicate conserved amino acids and dashes indicate missing amino acids. The termination of the small delta antigen is indicated by a block letter S followed by an arrow and a vertical bar. The coi-coil domain is doubly underlined. The nuclear localization signal is marked by a thin line. The RNA-binding domain is marked by a thick line. The large delta antigen package signal is marked by a hatched bar. Figures in parentheses indicate genotype. Sources of isolates are indicated: TWD62, our laboratory; J, Japan-1 (Imazeki et al., 1990); I, Italy (Wang et al., 1986); and P, Peru1 (Casey et al., 1993).

Fig. 4. Phylogenetic analysis of HDV isolates. Analysis of sequencing data was performed by maximum parsimony using PAUP (Swofford et al., 1996). Sequences of South American isolates, i.e. Peru1, Peru2, Colombia, US2, US3 and Japan3 (Casey et al., 1993), Lebanon (Lee et al., 1992), Nauru (Chao et al., 1991), China (Zhou et al., 1994), Italy (Wang et al., 1986), US1 (Makino et al., 1987), France (Saldanha et al., 1990), Japan-1 (Imazeki et al., 1990), T3 (Lee et al., 1996). The remaining isolates, beginning TW (Taiwan), are from our laboratory.
Table 3. Kimura’s two-parameter distances of the whole genomic sequences between a novel isolate (TWD62) and isolates of three HDV genotypes

Genotypes of each isolate are given in parentheses. Sources of isolates: Peru1, Casey et al., (1993); Italy, Wang et al. (1986); US, Makino et al. (1987); Japan-1, Imazeki et al. (1990); Taiwan3, Lee et al. (1996). The remaining isolates, beginning TW (Taiwan), are from our laboratory.

<table>
<thead>
<tr>
<th></th>
<th>Italy (I)</th>
<th>US (I)</th>
<th>TW2667 (I)</th>
<th>TW2476 (Ila)</th>
<th>Taiwan3 (Ila)</th>
<th>Japan-1 (Ila)</th>
<th>TWD62 (Iib)</th>
<th>Peru1 (III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy (I)</td>
<td>0.1092</td>
<td>0.1507</td>
<td>0.3138</td>
<td>0.3073</td>
<td>0.3063</td>
<td>0.3452</td>
<td>0.5390</td>
<td></td>
</tr>
<tr>
<td>US (I)</td>
<td></td>
<td>0.1479</td>
<td>0.3155</td>
<td>0.3051</td>
<td>0.3142</td>
<td>0.3618</td>
<td>0.5421</td>
<td></td>
</tr>
<tr>
<td>TW2667 (I)</td>
<td></td>
<td></td>
<td>0.3310</td>
<td>0.3339</td>
<td>0.3142</td>
<td>0.3976</td>
<td>0.5423</td>
<td></td>
</tr>
<tr>
<td>TW2476 (Ila)</td>
<td></td>
<td></td>
<td></td>
<td>0.0510</td>
<td>0.0586</td>
<td>0.2792</td>
<td>0.5555</td>
<td></td>
</tr>
<tr>
<td>Taiwan3 (Ila)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0613</td>
<td>0.2667</td>
<td>0.5459</td>
<td></td>
</tr>
<tr>
<td>Japan-1 (Ila)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2599</td>
<td>0.5287</td>
<td>0.5102</td>
</tr>
<tr>
<td>TWD62 (Iib)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peru1 (III)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Europe. Within the genotype I, there were two geographically classified subgroups: the Asian–Pacific subgroup and the Western subgroup. However, bootstrap values for each subgroup were below 70% and were not sufficient to support the concept of distinct subtypes. The Japan-1, the T3, and the eight isolates from our laboratory belong to the genotype IIa (previously defined as genotype II) (Casey et al., 1993). The isolates from five patients (TW1025, TWD19, TWD62, TW271 and TW247) appeared to form a novel monophyletic group supported by a bootstrap value of 100%. The reconstructed phylogeny of HDV suggests a close relationship between genotype IIa and the novel isolates; these two subgroups constituted a monophyletic group (with a bootstrap value of 90%) and formed a sister group to type I. MEGA analyses based on Kimura’s two-parameter distance matrix also identified a tree with an identical topology to the PAUP tree. Distances between the novel isolate and other genotypes were longer than those within any single genotype; however, the distances were shorter between the novel isolate and the genotype IIa than those between the novel isolate and other genotypes (Table 3). They are thus defined as genotype IIb. Phylogenetic analysis of HDV isolates based on whole genomic sequences or amino acid sequences resulted in similar evolutionary trees (not shown) to that based on the sequence from nt 911–1260 shown in Fig. 4. HDV genotyping based on phylogenetic analyses was completely identical to the results based on RFLP analysis.

Discussion

We recently developed a simple HDV genotyping method based on RFLP analysis (Wu et al., 1995b). Using this method, we found that HDV genotype II is not only the dominant strain in patients with acute hepatitis, but is also highly prevalent in the prostitutes from Taiwan, who are the most common source of infection (Wu et al., 1990). This study also indicates that the RFLP method is not only useful for the differentiation of the three known HDV genotypes, but is also of value for the discovery of novel HDV strains. It will thus be useful for a global molecular/epidemiological survey of the distribution of HDV genotypes. The novel strain we have identified was found in patients with acute hepatitis as well as in prostitutes, and its prevalence rate in the latter group was higher than that of genotype I. It is concluded that the new HDV strain does exist in Taiwan and that it may be transmitted via sexual contact and induced hepatitis.

There are several lines of evidence to support the idea that this novel HDV isolate is a distinct strain. First, it has a unique XhoI haplotype that is distinct from the three previously published genotypes. HDV genotype III does not have an XhoI cutting site within the amplified sequence from nt 889–1334, whereas genotype IIa (previously defined as genotype II) has two XhoI cutting sites within this region. Genotype I has only one of the two XhoI cutting sites of the genotype IIa and the novel strain (Iib) has only the other site within the amplified sequence. The association of XhoI haplotypes with HDV genotypes may have important evolutionary implications.

Second, the novel HDV group showed 14.6–32.6% divergence in nucleotide sequence in the nt 911–1260 region with the three currently known genotypes, whereas the divergence in the same region among isolates of the same genotypes was generally < 10% (Casey et al., 1996; Wu et al., 1995c; Lee et al., 1996). The divergence among isolates of this novel group was surprisingly low (< 3%), and was lower than that within the three known genotypes. This finding indicates that this novel group might have evolved much later than the three previously identified genotypes. Recently, a high divergence among Italian genotype I isolates was found; this is believed to result from an earlier and multiple introduction of this genotype to Italy (Niro et al., 1997). The study of HDV sequences of various genotypes in different areas of the world will be of great value to the understanding of the evolution and spread of HDV.
Finally, the strongest evidence comes from phylogenetic analyses based on multiple methods. Phylogenetic analysis using PAUP, MEGA or PHYLIP all revealed three major groups of HDV isolates (supported by bootstrap values > 99%) that are basically consistent with the previously proposed classification by Casey et al. (1993, 1996). In addition, a novel monophyletic group (supported by a 100% bootstrap value) was found. The novel group is closer to genotype Ila in evolutionary history than to other genotypes; it is thus defined as genotype IIIb.

Based on this analysis, both HDV genotypes I and II in Taiwan were closer to Japanese genotypes than to those from other areas in the world. It is reasonable to infer that both HDV genotypes I and II in Taiwan shared a common ancestor with the Japanese strains. It is surprising that the composition of HDV genotypes is so complex in Taiwan compared to other areas. This might be due to a high HBV carrier rate in this area, which favours maintenance and propagation of any introduced genotype, the geographic location, which is convenient for busy world trading and the influx of people from many countries, and also the fact that different populations from various countries have occupied or lived in this area.

There are variations in the divergence of nucleotide sequence in various regions of the HDV genome among the novel isolate and the three previously identified genotypes. The location of conserved regions important for virus replication are generally consistent with previous reports (Chen et al., 1992; Casey et al., 1993; Glenn et al., 1992; Lee et al., 1993, 1994; Lazinski & Taylor, 1993; Lai, 1995). The definition of highly conserved domains may help us to determine the precise locations of these domains, which may be of value in virological studies and the design of targets for antiviral therapy. Although both the N and C termini of the delta antigen vary greatly among different genotypes, they are rather conserved within the same genotype (Chao et al., 1991; Wu et al., 1995c; Lee et al., 1996). These regions appear to be genotype-specific and may be useful for the differentiation of HDV genotypes. Moreover, the divergence at the N and C termini between genotypes may influence the efficiency of HDV replication, packaging and the associated disease course. This hypothesis still awaits further studies using in vitro cell culture systems and animal models. There have been few reports comparing the various domains of the delta antigens among genotypes; this study of naturally occurring mutants may provide information for further virological studies.

Interestingly, the N and C termini of the delta antigens of the novel isolate showed a higher identity with those of genotype II, while the middle part (the nuclear localization signal and the RNA-binding sequences) showed a higher identity to genotype I. The mosaic composition of the delta antigen of this novel HDV strain suggests that it may have arisen from the recombination of HDV genotype I and II genomes during replication. Both HDV genotypes I and II are frequently found in Taiwan (Wu et al., 1995b, 1995c). Furthermore, we recently found that mixed infections of genotypes I and II exist in high-risk groups (prostitutes and their clients) (unpublished observations). These situations may provide an environment for recombination to occur during HDV replication. This hypothesis awaits further studies.

In summary, RFLP analysis is not only very useful for the initial genotyping of HDV in a large number of samples, but can also lead to the discovery of novel HDV strains. Sequencing and phylogenetic analysis have revealed that the novel isolate we have identified forms a novel monophyletic group (IIb) that is close to genotype Ila. The divergence in various regions of HDV genomes, among the novel isolate and different genotypes, may have important virological and clinical implications.

This study was supported by a grant NSC86-2314-B-010-051 from the National Science Council, Taiwan, Republic of China.

References


Received 27 August 1997; Accepted 19 December 1997