Characterization of hepatitis D virus genotype III among Yucpa Indians in Venezuela

Tatsunori Nakano,1,3 Craig N. Shapiro,1 Stephan C. Hadler,1 John L. Casey,2 Masashi Mizokami,3 Etsuro Orito3 and Betty H. Robertson1

1 Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
2 Georgetown University, Washington DC, USA
3 Second Department of Medicine, Nagoya City University Medical School, Nagoya 467-8601, Japan

The complete genome sequences of hepatitis D virus (HDV) strains isolated from three Yucpa Amerindians in Venezuela were determined and found to be genotype III. Comparison of these three genotype III sequences demonstrated the presence of a hypervariable region containing numerous substitutions, insertions/deletions and a highly conserved region containing the self-cleavage domains, which have been reported previously for genotypes I and II. Amino acid changes within the first 90 amino acids of the hepatitis D antigen (HDAg) were found in the genotype III sequences, while the remainder of the HDAg-coding sequence was conserved. The secondary structure for the RNA-editing site differed between genotypes I and III. It was concluded that the serious delta hepatitis outbreaks characterized epidemiologically in the Yucpa Amerindians were caused by HDV genotype III isolates that were related to HDV genotype III isolates from other regions of South America.

Hepatitis D virus (HDV) is a subviral agent that needs the presence of hepatitis B virus (HBV) to be infectious. Because of an increased risk of fulminant hepatitis and progression to serious liver disease, HDV super- or co-infection is a major health risk for persons with chronic and acute HBV infection and for populations that have a high carrier rate of HBV. HDV is composed of a 1.7 kb covalently closed, circular, negative-sense RNA surrounded by a hepatitis D antigen (HDAg) capsid and a lipid bilayer containing the HBV surface antigen (Rizzetto et al., 1980; Bonino et al., 1986; Makino et al., 1987).

A single open reading frame for the HDAg is contained on the positive-sense strand of HDV RNA (Wang et al., 1986).

Three phylogenetically distinct genotypes (I, II and III) of HDV that have different geographical distributions have been identified. Genotype I is found in North America, Europe, Africa, east and west Asia and the South Pacific (Zhang et al., 1996; Shakil et al., 1997; Niro et al., 1997). Genotype II has been found only in east Asia (Imazeki et al., 1991; Wu et al., 1995, 1998; Lee et al., 1996; Sakugawa et al., 1999), while genotype III has been identified only among cases in South America (Casey et al., 1993, 1996).

There have been reports of severe or fulminant hepatitis with unique clinical and morphological features in the northern regions of South America as far back as the 1930s. These infections, commonly known as Santa Marta hepatitis or Labrea hepatitis, have been identified as HBV and HDV super- or co-infections (Ljunggren et al., 1985; Buitrago et al., 1986a, b; Bensabath et al., 1987; Colichon et al., 1988; Gayotto, 1991). Hepatitis outbreaks with these clinical features and serological markers occurred in an epidemic identified in the Yucpa Amerindians in the northwestern portion of Venezuela between 1979 and 1981 (Hadler et al., 1984). In this study, the complete HDV genome isolated from three individuals who developed hepatitis during this outbreak was sequenced and evaluated.

Samples from the three individuals, who were residents of three distinct but closely located villages, were tested and found to be positive for HBV and HDV markers, as described previously (Hadler et al., 1992; Fields et al., 1986). Samples were collected in 1990, although the infections occurred in the early 1980s. Nucleic acid was extracted from 50 µl of serum using the MasterPure RNA Purification kit (Epicentre). cDNA was prepared by reverse transcription using random hexamers (Boehringer Mannheim) and Moloney murine leukaemia virus reverse transcriptase (Boehringer Mannheim). One-tenth of the cDNA sample was amplified by PCR using Taq polymerase (Boehringer Mannheim). The conditions for PCR were 2 min at 95 °C, 45 cycles of 1 min at 94 °C, 1 min at 55 °C and 3 min at 72 °C and 1 cycle of 7 min

Author for correspondence: Betty H. Robertson.
Fax +1 404 639 1563. e-mail bjr1@cdc.gov

The GenBank accession numbers of the sequences reported in this paper are AB037947–AB037949.
Fig. 1. (a) Phylogenetic tree based on the complete sequences of HDV isolates. (b) Phylogenetic tree based on the sequence of nucleotide positions 911–1260 (numbered according to the Italy isolate). The isolates that have VnzD as the first four letters are the Venezuelan sequences determined in this study. Bootstrap values from 1000 replicates (Felsenstein, 1985) are shown for the respective branches. Peru-1 (L22063), Peru-2 (L22064) and Colombia (L22061) are used as representative isolates of genotype III. TWD62 (AF018077), isolate 1-18 (AF015442), isolate 2-05 (AF015443), isolate 3-25 (AF015444), isolate 4-15 (AF015445), isolate 5-01 (AF015446) and isolate 6-21 (AF015447) are used as representative isolates of genotype IIb. Japan-1 (X60193) and Taiwan-3 (U19598) are used as representative isolates of genotype IIa. Somalia (U81988), Lebanon-1 (M84917), Nauru (M58629), Taiwan-1 (M92448), China (X77627), Japan-3 (L22062), Japan-2 (D90190), US-1 (M28267), US-3 (L22065), France (D01075), Patient A (X85253), US-2 (L22066), Canada (AF098261), Central Africa (A000558) and Italy (X04451) are used as representative isolates of genotype I. Scale bar, substitution per site. (c) Graphic illustration of sequence variation and number of gaps needed for optimal alignment at each
at 72 °C. Amplicons for sequencing the entire HDV genome consisted of five overlapping fragments. Fragments were amplified using the primer pairs 308P (5′ TCCAGGACCCTTGGGGCAACA 3′) and 720N (5′ CTGGATCATGTTGCCAGCCCG 3′) for nucleotide positions 307–737 (nucleotide positions are according to the Italy isolate; Wang et al., 1986); 688P (5′ TGGCCCGCATGGGGCCCAACG 3′) and 889N (5′ TTCTTCTGGGTCCGGATGGAT 3′) for nucleotide positions 685–909; 853P (5′ CGGATGCCCAGTCGGACC 3′) and 1267N (5′ GAGGGAAGCCCTTGGGAACAAAGA 3′) for nucleotide positions 855–1287; 1267P (5′ TTCTTGTTCAGGGCTCTCTTC 3′) and 503N (5′ CCCCCGATAAGCCTCAGTGC 3′) for nucleotide positions 1264–485; and 887P (5′ AGATCCATGGCCGACCCGAAGAG 3′) and 1360N (5′ GGGCGAGAGGACATGGAGATTG 3′) for nucleotide positions 883–1374.

Using the PCR primers above, PCR products were sequenced in the presence of rhodamine terminators using a 377A DNA Sequencer (Applied Biosystems). An internal primer pair, 1660P (5′ AGCTTCGAGCTCCTCCCG 3′) and 60N (5′ ATTCGGCTCTTTTCCTCC 3′), was also used to sequence the fragment spanning nucleotide positions 1264–485. The Wisconsin package, version 10.1 (Genetics Computer Group, Madison, WI, USA), and the ODEN computer program, version 1.1.1 (Iwa, 1994), were used for sequence analyses.

The three complete HDV genomes were 1673 (VnzD8349), 1672 (VnzD8375) and 1674 (VnzD8624) nucleotides in length and were closely related to each other (97.0–98.5% nucleotide identity). The percentage identity to the other complete genotype III isolate, Peru-1, was 89–89.6% (Casey et al., 1993), while the identity to genotype I and II sequences was only 63.8–67.1% (Wang et al., 1986; Imazeki et al., 1991; Wu et al., 1998). Phylogenetic analysis of the Venezuelan HDV full-length genome sequences and other available full-length genome sequences was performed. Initial alignments were made using the GCG Pile-Up program. However, when we reviewed the computer-generated alignments, we noted multiple regions of apparent identity in the hypervariable region (Lee et al., 1996) that were not computer-aligned; therefore, alignments were manually corrected to produce optimal matching of the bases. Intermittent conserved segments were aligned first, followed by alignment of the sequences between the conserved segments, which were aligned for highest homology, inserting gaps if needed. The phylogenetic tree was constructed using the neighbour-joining method (Saitou & Nei, 1987), with genetic distances calculated using the six-parameter method (Gojobori et al., 1982). The tree showed that the three Venezuelan isolates clustered with the previously reported genotype III isolate, Peru-1, and were distinctly distant from the isolates reported as genotypes I or II (Fig. 1a). A more inclusive phylogenetic analysis using 30 partial HDV sequences (nucleotide positions 911–1260) was also performed (Fig. 1b). Once again, the Venezuelan sequences clustered with the two Peruvian and one Colombian isolate, confirming that the three Venezuelan isolates belong to genotype III.

An illustration of the sequence variation and number of gaps needed for alignment at each nucleotide position among genotype III sequences (the three Venezuelan sequences and Peru-1) is shown in Fig. 1 (i, ii); an analogous evaluation of the genotype I sequences from Fig. 1 (i) is shown in Fig. 1 (c, ii). The hypervariable region, which has been identified among genotype I sequences (Lee et al., 1996) and has multiple insertions/deletions and numerous nucleotide substitutions, was also observed among genotype III sequences. Sequences outside the hypervariable region were relatively conserved, with only 3.3% of the nucleotide positions containing changes. The genomic and anti-genomic self-cleaving sequence elements (Perrotta & Been, 1991) were well-conserved among genotype III as well as genotype I sequences. These elements, which are critical for HDV replication, were also well-conserved among all of the other sequences compared (data not shown). Two sequences, VnzD8439 and VnzD8624, had ambiguities in several nucleotide positions throughout the full-length genome (data not shown). This microheterogeneity in a single host has also been observed in genotype I and II infections (Chao et al., 1994; Lee et al., 1996).

When the translated amino acid sequences of the Venezuelan isolates were compared to each other and to the genotype III Peru-1 isolate, amino acid identity within HDAG among the genotype III Venezuelan isolates was 94.0–98.1%, and between the Venezuelan and the Peru-1 isolates, amino acid identity was 87.0–89.3%. Each Venezuelan sequence had unique amino acid substitutions, the majority (11/13) of which were clustered within the first 90 amino acids (Fig. 2a and b, i). This amino acid pattern was also seen when the three Venezuelan isolates were compared with the other available genotype III sequences (Fig. 2a and b, ii). Amino acid variability may reflect the fact that genotype III-specific amino acids within the amino terminal region of small-HDAG (S-HDAG) are not critical for replication of genotype III RNA (Casey & Gerin, 1998). However, the hydrophobic amino acids at positions 44, 51 and 58, which are thought to be important for
HDAg dimerization (Xia et al., 1992; Rozzelle et al., 1995), were all conserved; basic amino acids in the nuclear localization signal (NLS) (Xia et al., 1992) were also conserved among genotype III sequences.

In contrast, amino acids encoding the middle and carboxyl-terminal region of S-HDAg and the carboxy-terminal extension of large-HDAg (L-HDAg) were highly conserved within the genotype III isolates. Although we detected no evidence for the production of L-HDAg based upon the genotype III sequences available, sequence conservation of this region may indicate a functional role for this region of genotype III HDAg. The amino acid sequence of the terminal extension of L-HDAg has been reported to be different among different genotypes and to be conserved within genotypes I (Shakil et al., 1997) and II (Lee et al., 1996; Wu et al., 1998). Our data also suggest a similar conservation pattern among genotype III sequences. However, to confirm a pattern of conservation, additional sequence information from full-length genome sequences would be needed.

Fig. 2. (a) HDAg amino acid sequences of the Venezuelan isolates compared with other genotype III sequences. Dashes indicate amino acid identity with the consensus sequence. Amino acids for which a consensus sequences cannot be determined are indicated with an X. Asterisks indicate stop codons. The four Xs in VnzD8349 indicate the residues A or S, S or P, R or Q and D or A, respectively; the single X in VnzD8624 indicates an A residue or an E residue. Hashes indicate amino acid positions that are thought to be important for HDAg dimerization.

(b) Amino acid sequence variation at each amino acid position among (i) the three Venezuelan sequences and (ii) all available genotype III sequences. Sequence variation among the three Venezuelan sequences and Peru-1 is indicated between amino acid positions 1 and 113 and the sequence variation among the available six genotype III sequences is indicated between amino acid positions 114 and 214. Variation is calculated as described for Fig. 1(c). The domains involving HDAg dimerization, NLS and RNA binding, a domain which contains arginine-rich motifs (ARM) (Lee et al., 1993), and the terminal extension of L-HDAg are shown.
HDV genotype III in Venezuela

Fig. 3. (a) Nucleotide sequences of the RNA-editing target for genotype I sequences, which is formed between anti-genome RNA surrounding the edited A residue (nucleotide 1012), and nucleotide sequences around nucleotide 580 on the opposite side of the unbranched rod structure of HDV. Nucleotide positions are numbered according to the Italy isolate. (b) Nucleotide sequences of the RNA-editing target for genotype III sequences available. The RNA-editing position is located at nucleotide position 1014. Nucleotide positions are numbered according to the Peru-1 isolate. Dashes indicate identity to the consensus sequence. Dots indicate gaps needed to form an unbranched rod structure. (c) A particular base-paired structure formed by genotype I sequences for RNA editing. (d) The predicted similar structure formed by the genotype III sequences for RNA editing. Solid lines between the nucleotide sequences indicate base pairing.

L-HDAg originates from S-HDAg as a result of RNA editing. RNA editing at nucleotide 1012 replaces the stop codon (UAG) of S-HDAg with a tryptophan codon (UGG), resulting in the translation of 19–20 additional amino acids to generate L-HDAg (Luo et al., 1990; Wang et al., 1992; Casey et al., 1992). Genotype I sequences have a mixture of sequences with either an A residue or a G residue at the RNA-editing site (Fig. 3a, nucleotide 1012). The six genotype III sequences available for analysis all have an A residue at the RNA-editing site (Fig. 3b, nucleotide 1014). It has been shown for genotype I that a particular base-paired structure is formed between nucleotides surrounding the edited A residue and those around nucleotide 580, which is on the opposite side of the branched rod structure typical of HDV RNA (Fig. 3c) (Casey et al., 1992; Casey & Gerin, 1995). The structure formed by genotype III RNA is significantly different from the genotype I RNA structure in that there is a bulge of five nucleotides, including the A residue, in the region of sequence to be edited (Fig. 3d). This deviation, which is conserved in all of the additional genotype III sequences reported here (Fig. 3b), is likely to be significant for the editing activity of genotype III RNA, as site-directed mutagenesis has shown that editing of the HDV genotype I RNA is extremely sensitive to the identity of the base opposite the A residue (Casey et al., 1992; Casey & Gerin, 1995). This deviation could also result in less RNA editing, thereby limiting the amount of L-HDAg.
produced and could influence the pathogenesis of HDV infection (Govindarajan et al., 1993; Tang et al., 1994; Yang et al., 1995): L-HDAg is postulated to reduce genomic replication and limit infection (Glen & White, 1991; Chao et al., 1990). This might be reflected in genotype III infections, which appear to cause a more severe form of HDV infection (Casey et al., 1993, 1996). Additional information on the sequences associated with RNA editing and the different HDV genotypes is needed to conclusively identify the relationship between HDV RNA editing and severe liver disease.

In the early 1980s, the Yucpa Amerindians in Venezuela experienced a devastating HDV epidemic, with high rates of fulminant hepatitis and progressive chronic hepatitis in those who developed chronic HDV infection (Hadler et al., 1992). Our studies confirm that the HDV genotype responsible for this outbreak of severe delta hepatitis was genotype III, the same genotype identified among delta hepatitis cases in other northern regions of South America, specifically Colombia and Peru, and different from genotypes originating from other parts of the world (Casey et al., 1993, 1996). We have previously reported that the HBV genotype infecting the Yucpa delta hepatitis patients is genotype F (Nakano et al., 2001), which also appears to be geographically and ethnically focused among Amerindians in South and Central America (Norder et al., 1993; Arauz et al., 1997). The genetic difference between this delta hepatitis agent and other genotypes suggests that the South American HDV isolate is as different from other HDV as the South American genotype F HBV isolate is from other human HBV. The observations that genotype F HBV and genotype III HDV sequences are common among isolated indigenous Amerindians in northern regions of South America and that they are distantly related to those found in other areas of the world suggest that cocirculation of genotype III HDV and genotype F HBV is not coincidental (Casey et al., 1996). These data further indicate that the HDV genotype prevalent in these populations was not introduced by a non-South American source. The severe liver disease observed in these infections might reflect the coevolution of these two genetic variants, although a unique virus–host interaction in these genetically isolated populations cannot be excluded.

References


Received 2 April 2001; Accepted 4 June 2001