Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia

Valeria Ivaniushina,1,3 Nadjia Radjef,3 Marfa Alexeeva,2 Elyanne Gault,3 Sergei Semenov,2 Mohammed Salhi,3 Oleg Kiselev1 and Paul Dény3

1Influenza Institute, St Petersburg, Russia
2Yakutsk City Hospital, Yakutsk, Russia
3Laboratoire de Bactériologie, Virologie – Hygiène, Equipe d’Accueil ‘Agents Transmissibles et Hôtes, Signalisation Cellulaire, Oncogénèse’, Hôpital Avicenne, UFR Santé Médecine Biologie Humaine, Université Paris 13, 125 route de Stalingrad, 93009, Bobigny cedex, France

Currently, three genotypes of hepatitis delta virus (HDV) are described. The most common, genotype I, has a worldwide distribution; in contrast, genotype II has been found previously only in Japan and Taiwan, while genotype III is found exclusively in South America. Considering the high prevalence of HDV in Northern Siberia (Russia), restriction fragment length polymorphism (RFLP) was used to analyse HDV genotypes from 29 infected patients living in Yakutia. Of these isolates, 11 were characterized by partial nucleotide sequencing and two isolates were completely sequenced. Phylogenetic inference methods included maximum parsimony, maximum likelihood and distance analyses. A restriction pattern consistent with HDV genotype I was found in 14 samples, while the remaining 15 showed a different restriction pattern, inconsistent with any known genotype. Five Yakutian HDV isolates with the type I restriction pattern were sequenced and confirmed to be affiliated with genotype I, although the phylogenetic results indicate that they were heterogeneous and did not cluster together. Sequencing of eight isolates with the new RFLP pattern revealed that these isolates were most closely related to HDV genotype II. In contrast to HDV Yakutian genotype I sequences, all of these type II sequences formed a well-defined clade on phylogenetic trees. Comparison of clinical presentations during hospitalization between patients infected with HDV type I (n = 14) and type II (n = 15) did not reveal any differences in the severity of infection. These data indicate that the distribution of genotype II is not restricted to Taiwan or Japan, but spreads over Northern Asia, appearing in the native population of Yakutia. Type II Yakutian strains appeared to form a well-defined subclade and could be associated with severe chronic hepatitis in this area.

Introduction

In some regions of Russia and the former Soviet Union, a high prevalence of hepatitis delta virus (HDV) infection has been reported (Favorov et al., 1997; Flodgren et al., 2000); however, there has been only limited genetic information about HDV isolates. Comparison of published HDV cDNA sequences led to the identification of three distinct HDV genotypes, each of them with a different geographical distribution and pattern of pathogenesis (Casey et al., 1993; Wu et al., 1998). In contrast to the most common genotype of HDV, genotype I, which is widely distributed and found all over the world, HDV genotype II has been identified only in Japan and Taiwan (Imazeki et al., 1990; Lee et al., 1996). It has been suggested that genotype II infection causes a less serious form of disease than infection with genotype I originating from the same area (Wu et al., 1995; Sakugawa et al., 1999). Genotype III, found exclusively in the northern region of South America, is associated with outbreaks of severe fulminant hepatitis (Casey et al., 1993, 1996).

In Yakutia (Sakha Republic, Russia), a region that encompasses vast areas of Central and Northern Siberia, hepatitis B virus (HBV) infection is a serious health care problem. Indeed,
HBV prevalence in Yakutia is 3–4 times higher than the average for Russia. Furthermore, HDV markers are detectable in 18–20% of HBV surface antigen (HBsAg)-positive hepatitis cases (Viazov et al., 1989; Alexeeva et al., 1998), indicating a high level of HDV endemicity in Yakutia. To assess the genetic variations of HDV genomes from this area, we first screened 29 samples from patients who were chronically infected with HDV by restriction fragment length polymorphism (RFLP) analysis of amplified HDV cDNAs. We then focused on 13 specific isolates for sequence and phylogenetic analyses. The complete nucleotide sequences of two isolates, which presented an original RFLP pattern, were obtained. The results indicate that two HDV genotypes (I and II) coexist in this area. Yakutian genotype II isolates form a distinct cluster on HDV phylogenetic trees, possibly representing a specific genotype II subclade.

**Methods**

- **Patients.** Sera were obtained from 29 patients with chronic delta hepatitis. Patients were admitted to Yakutsk City Hospital (Yakutia, Russia) between 1996 and 1998. Among them, 25 (86.2%) were ethnic Yakuts and 4 (13.7%) were ethnic Russians. None of the patients was directly related to each other. Of these 29 patients, 20 lived in Yakutsk, the main city of the region, while the others came from remote towns and villages. All of the patients had a history of chronic liver disease and were positive both for HBsAg (Orgenics) and for antibody to the hepatitis delta antigen (HDAg) (Akvapast). History of past acute hepatitis was described for 12 of the patients. An HBV-positive in-house contact existed for five patients (plus HDV in two cases); previous times of hospitalization, including those for parenteral procedures, were seen for 10 patients; three infected individuals were health care workers (one surgeon and two nurses) and one patient had used drugs previously. No risk factors were evident for the remaining patients. No human immunodeficiency virus or hepatitis C virus infection was evident, as determined through antibody testing. The gravity and prognosis of chronic hepatitis were evaluated according to the classification that was defined at the International Congress in Los Angeles, USA, in 1994.

- **RT–PCR and RFLP analysis.** Sera from all 29 patients were screened for HDV RNA by RT–PCR and analysed by RFLP. HDV RNA was extracted from 100 µl serum by the guanidine–chloride method and dissolved in 50 µl DEPC-treated water. For reverse transcription, 5 µl extracted RNA was mixed with random primers (0.4 µM) and dNTP (0.5 mM) and denatured for 3 min at 95°C. Denatured RNA was added to the reverse transcription mixture (total volume 25 µl) containing 20 U RNasin (Promega) and 100 U SuperScript reverse transcriptase (Life Technologies) in the buffer supplied by the manufacturer. The reaction was incubated at 42°C for 45 min and stopped by incubation at 94°C for 5 min. The PCR reaction mixture (total volume 40 µl), containing 0.25 pmol/µl of primers and 1 U AmpliTaq Gold polymerase (PE Applied Biosystems) in the buffer supplied by the manufacturer, was added to 10 µl of the reverse transcription reaction and covered with 50 µl of mineral oil. PCR was carried out in a thermocycler (Gene Amp PCR System 2400) (PE Applied Biosystems) under the following conditions: 9 min at 94°C, followed by 40 cycles of 45 s at 94°C, 30 s at 58°C and 45 s at 72°C with a final extension step of 5 min at 72°C. Positive and negative controls were included in each set of reactions. Strict procedures were followed to avoid false-positive results. Results of PCR were analysed by electrophoresis in 1.3% agarose gels. For RFLP analysis, products of amplification with primers 900s and 1280as (Fig. 1) from 50 µl reaction were extracted with phenol–chloroform, precipitated with ethanol and dissolved in 8 µl water. Digestion was performed in 10 µl samples with 5 U Smal (Promega) for 2 h at 25°C. Digested products were analysed by electrophoresis in 2% agarose gels.

- **Sequence strategy.** Two overlapping regions of the HDV genome, R0 and R1, were sequenced initially (Fig. 1). Region R0, which represents the 3'-terminal part of the HD gene and is flanked with primers 900s and 1280as, was proposed by Casey et al. (1993) for HDV genotype classification and has been used extensively since then for phylogenetic analyses. R1 is a genomic region flanked with primers 320s and 1280as; this region was expected to give stronger phylogenetic information as it is 2.5 times longer than R0 and also includes more regions of variability (Wu et al., 1998). In order to obtain complete HDV sequences, additional primers were used. Primers 900s and 320as amplified fragment R2 (about 1100 bp), which covered the whole HDAg protein-coding sequence (Fig. 1). After sequencing the R1 and R2 regions of two HDV isolates, a new pair of primers (1480s and 440as) was designed based on sequence information obtained in order to verify region R3, which encompasses primers 320s and 320as. Due to considerable overlapping of the R1, R2 and R3 regions, most parts of the HDV genome were sequenced repeatedly. Two independent clones were sequenced for each of the R1, R2 and R3 regions for both fully sequenced Yakutian HDV genomes (p126 and p1262). PCR products were sequenced either directly or after cloning into the pCR2–TA cloning vector (Invitrogen). For direct sequencing, PCR products were gel-purified through mini-columns (Centricon Micro-Concentrators 50) (Millipore). Purified DNA fragments were sequenced directly on both strands on an

---

**Table 1.** Strategies for Yakutian HDV genome amplification. Regions R0 (primers 900s–1280as), R1 (320s–1280as), R2 (900s–320as) and R3 (1480s–440as) are indicated by arrows representing primer locations. Note that region R2 encompasses the whole HD protein open reading frame located on the antigenomic strand. The sequence and position of primers according to the numeration of Wang et al. (1986) are listed below.

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Sequence 5'→3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>320s</td>
<td>305-327</td>
<td>ccagaggacccctcagcgaac</td>
</tr>
<tr>
<td>320as</td>
<td>327-306</td>
<td>gtgcgtgaaggggtcctctg</td>
</tr>
<tr>
<td>440as</td>
<td>432-452</td>
<td>aacatcactgcgtacgccc</td>
</tr>
<tr>
<td>900s</td>
<td>885-908</td>
<td>catgccccagcagagagaaag</td>
</tr>
<tr>
<td>1280as</td>
<td>1285-1261</td>
<td>gaggagaagccctcgaacaaaga</td>
</tr>
<tr>
<td>1480s</td>
<td>1462-1486</td>
<td>tcaatccgagtccttattgtctg</td>
</tr>
</tbody>
</table>

**Fig. 1.** Strategy for Yakutian HDV genome amplification. Regions R0 (primers 900s–1280as), R1 (320s–1280as), R2 (900s–320as) and R3 (1480s–440as) are indicated by arrows representing primer locations.
Table 1. SmaI digestion of the R0 fragment (405 bp) for reference and Yakutian isolate sequences

<table>
<thead>
<tr>
<th>Samples</th>
<th>Size (bp) after SmaI digestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype I</td>
<td>227, 178</td>
</tr>
<tr>
<td>Genotype II</td>
<td>405</td>
</tr>
<tr>
<td>Genotype III</td>
<td>301, 104</td>
</tr>
<tr>
<td>Yakutian samples (n = 14)</td>
<td>227, 178</td>
</tr>
<tr>
<td>Yakutian samples (n = 15)</td>
<td>178, 117, 110</td>
</tr>
</tbody>
</table>


**Secondary structure determination.** Secondary structure determination was performed using the complete genome and antigenome sequences of HDV genotypes I (A20; Italy), IIA (Japan), IIB (Taiwan-TW2b) and III (Peru1) and the Yakutian sequences pt26 and pt62. The mfold program, version 3.1 (Mathews et al., 1999; Zuker et al., 1999), predicts the possible secondary structures for RNA sequences. This program was made available by M. Zuker (http://bioinfo.math.rpi.edu/~zukerm/).

**Phylogenetic analysis.** Editing of raw sequence data was carried out with Sequence Navigator, version 1.0.1 (PE Applied Biosystems). Sequences were aligned with CLUSTAL W 1.8 (Thompson et al., 1994) using the ‘fast’ option and different levels of gap extension penalty (2, 5, 10). Minimal manual corrections were performed with SeqPup. Neighbour-joining (NJ) analyses of the sequences were carried out using the DNADIST and NEIGHBOR programs of the PHYLIP package, version 3.572 (Felsenstein, 1989), or PAUP 4.0b3 (Swofford, 1998) with pairwise distances estimated through Kimura 2-parameter or maximum-likelihood (ML) options. Maximum parsimony (MP) analyses were carried out using PAUP 4.0b3 with branch-and-bound (for less than 20 taxa) or heuristic searches. All characters were weighted equally and tested as unordered. ML analyses were performed using PAUP 4.0b3 with base frequencies, proportion of invariable sites, shape parameter for gamma distribution of variable sites and substitution rate-matrix estimated through ML analyses from the data. Due to computational limits imposed by the ML algorithms, the number of sequences analysed was limited, as described in Results, and bootstrapping was performed on a limited number of replicates (n = 100). Otherwise, we analysed the robustness of different branches by bootstrapping (10²–10⁴ replicates). All trees were visualized with TreeView (Page, 1996).

### Results

**RFLP and sequence analyses of Yakutian isolates**

In order to assess the genetic variability of HDV from a highly endemic area, samples from 29 HDV-viraemic patients originating from Yakutia were analysed. Among them, 14 samples (48.3%) showed a SmaI restriction pattern advocating a HDV genotype I profile, while the remaining 15 had a different restriction pattern that was inconsistent with any of the known genotypes (Table 1). Thirteen samples (five with a genotype I restriction pattern and eight with the uncharacterized pattern) were selected for further analysis by sequencing. Specifically, all patients from distant villages and towns who were suspected to be infected with isolates originating from geographically distinct locations were included; furthermore, patients representing different patterns of illness (asymptomatic, benign and severe) were selected for HDV nucleotide sequence analyses (Table 2).

HDV R0 and R1 nucleotide sequences (see Methods) were obtained from 8 and 11 isolates, respectively. Some isolates were analysed in both regions with similar results. Furthermore, the whole genomic sequence from two isolates (pt26 and pt62) was characterized. No identical virus sequences were found among the 13 HDV-infected patients, confirming the absence of PCR contamination and allowing phenetic and phylogenetic analyses.

Table 2. Patients included in HDV sequence analysis

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Sex</th>
<th>Liver disease (status)</th>
<th>Origin</th>
<th>Ethnic origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt6</td>
<td>27</td>
<td>M</td>
<td>Cirrhosis (died)</td>
<td>Yakutsk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt8</td>
<td>39</td>
<td>F</td>
<td>Chronic hepatitis (benign)</td>
<td>Tattinskii ulus</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt12</td>
<td>17</td>
<td>M</td>
<td>Cirrhosis</td>
<td>Yakutsk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt13</td>
<td>17</td>
<td>M</td>
<td>Chronic hepatitis (grave)</td>
<td>Yakutsk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt26</td>
<td>35</td>
<td>M</td>
<td>Cirrhosis</td>
<td>Nurbinskii ulus</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt29</td>
<td>19</td>
<td>F</td>
<td>Chronic hepatitis (grave)</td>
<td>Yakutsk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt30</td>
<td>20</td>
<td>M</td>
<td>Cirrhosis</td>
<td>Tattinskii ulus</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt51</td>
<td>52</td>
<td>F</td>
<td>Cirrhosis (died)</td>
<td>Yakutsk</td>
<td>Russian</td>
</tr>
<tr>
<td>Pt62</td>
<td>19</td>
<td>M</td>
<td>Chronic hepatitis (grave)</td>
<td>Verhneviuisk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt63</td>
<td>42</td>
<td>M</td>
<td>Chronic hepatitis (grave)</td>
<td>Verhneviuisk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt1245</td>
<td>55</td>
<td>M</td>
<td>Chronic hepatitis (grave)</td>
<td>Abyiskii ulus</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt704</td>
<td>38</td>
<td>F</td>
<td>Chronic hepatitis (asymptomatic)</td>
<td>Yakutsk</td>
<td>Yakut</td>
</tr>
<tr>
<td>Pt724</td>
<td>21</td>
<td>M</td>
<td>Chronic hepatitis (grave)</td>
<td>Yakutsk</td>
<td>Russian</td>
</tr>
</tbody>
</table>
Fig. 2. Unrooted trees obtained from distance matrix analyses of three different genomic regions from Yakutian isolates and reference sequences. (a) R0 (353 bp), (b) R1 (936 bp) and (c) complete genome regions are shown. The three representations indicate distances estimated through ML with NJ reconstruction. When > 60%, the BVs for 104 replicates are indicated in italic typeface. Scale represents the number of nucleotide substitutions per site. Reference sequences for phylogenetic analysis were chosen among partial and complete HDV sequences deposited in the GenBank/EMBL/DDBJ databases under the following accession numbers: Cagliari (X85253), Canada (AF098261), China (X77627), Colombia (L22061), Ethiopia (U81988), Chimpanzee A20 (Italy) (X04451), Japan (X60193), Lebanon (M84917), Nauru (M58629), Peru1 (L22063), Peru2 (L22064), Somalia (U81988), Taiwan (M92448), Taiwan3 (U19598), Taiwan-TW2b (AF018077), US-1 (D01075), US-2 (L22066), Woodchuck 15 (W15) (M21012) and Woodchuck 5 (W5) (AJ307077). Yakutian sequences obtained through this study correspond to EMBL accession numbers AJ309868–AJ309881.

The estimation of nucleotide sequence similarity among the Yakutian isolates showed that they form two distinct groups, in accordance with the RFLP results. Given the R1 genomic region, the first group was more closely related to the reference genotype I sequences than to the genotype II and III sequences. Yakutian type I isolates had inter-strain genetic distances \(d\), estimated through Kimura 2-parameter analysis, ranging from 7.4 to 11.2%. Their genetic distances from the reference strain A20 range between 8.4 and 12.3%. Inside the group with the new restriction pattern, virus sequences were less
Phylogenetic analyses

Phylogenetic analyses were performed on three sets of data. For R0, R1 and complete HDV genomes (see Methods), each set included 19, 15 and 16 reference sequences from databases and 8, 11 and 2 sequences obtained from Yakutian isolates, respectively. For these three regions, results of distance analyses followed by NJ tree reconstructions (distances estimated through ML) are shown in Fig. 2 as radial representations. NJ tree reconstructions resulted in phylogenetic trees of similar topology. Genotype I, II and III sequences defined previously are easily distinguishable. Based on analysis of regions R0 and R1 (Fig. 2a, b), five samples (pt8, pt12, pt30, pt51 and pt724) with the type I Smal restriction profile were scattered among type I sequences. As proposed previously, genotype II appears to be subdivided into subtype IIA and Taiwan-TW2b (Wu et al., 1998). Type IIA corresponds to type II described previously and includes the Japan (Imazeki et al., 1991) and Taiwan3 sequences (Lee et al., 1996). All Yakutian HDV sequences with the new original restriction pattern (pt6, pt13, pt26, pt29, pt62, pt63, pt245 and pt704) form a monophyletic group on each data set analysed, the branch of which arises between Taiwan-TW2b and type IIA nodes (Fig. 2a, b). Furthermore, when the complete sequences are analysed (Fig. 2c), pt26 and pt62 form, together with Japan and Taiwan3, a distinct group that excludes Taiwan-TW2b. These analyses indicate that Yakutian isolates consist of two distinct profiles corresponding to two different genotypes. Yakutian type I sequences are distantly related to each other and scatter among the different type I reference sequences studied, while Yakutian type II sequences form a more closely related, well-defined subgroup on tree reconstructions.

To study these results further, the sequence data sets were analysed using MP with heuristic or branch-and-bound searches. A representative tree obtained after a $10^4$ replicate bootstrap analysis of the R1 region is shown in Fig. 3. It indicates again that, in contrast to genotype I sequences where no strong specific subgroups were observed (with the exception of Taiwan and pt8), Yakutian genotype II sequences (pt6, pt13, pt26, pt29, pt62, pt63, pt245 and pt704) formed a monophyletic clade, irrespective of the data set being studied and the kind of analysis made. Furthermore, these results suggest that these Yakutian isolates form a common clade together with type IIA sequences. Similar tree topologies were also obtained for the R0 region using MP analyses (data not shown). The confidence of this clade was appraised through bootstrap analyses using $10^3$–$10^4$ replicates (Figs 2 and 3). It was demonstrated also that the monophyly of Japanese, Taiwanese (Taiwan3 isolate) and the novel Yakutian isolates was reproducible for every partial or complete sequence studied from different data sets, [MP/NJ bootstrap value (BV)] for R0 = 99.6–99.9, R1 = 99.5–99.8, complete sequence = 100/100]. These results strongly suggest that a common ancestor existed for these Taiwan, Japanese and Yakutian type II isolates.

Complete cDNA sequence and deduced HD protein amino acid sequence of two Yakutian isolates

The length of complete HDV sequences deposited in databases varies from 1672 (Canada) to 1683 (USA) nt, of which most isolates are between 1676 and 1679 nt in length. The length of the genome is 1685 nt for pt26 and 1688 nt for pt62, making Yakutian sequences the longest HDV genomes sequenced to date. When the alignment of two complete HDV cDNA sequences from Yakutia (pt26 and pt62) with prototype sequences of genotypes I, IIA, IIB and III was performed (data not shown), both Yakutian genomes were found to contain
Fig. 4. Secondary structure of HDV antigenomic RNA obtained using the mfold program on (a) complete HDV sequence and (b) focused on the editing site and the 3’ end of the LHD gene. The two stops codons in phase are indicated in their respective position in bold typeface. Note the difference of the predicted secondary structures around the editing region between type I and II sequences.

signals complementary to the start codon and the two stop codons for the small HD (sHD) and the large HD (LHD) proteins. However, a mutation in the RNA-editing site (U → C, position 1019 in pt26 sequence) was not detected in the sequenced clones obtained from serum samples. As demonstrated previously (Chao et al., 1991), alignment analysis indicated conserved areas corresponding to sequences surrounding the genomic and antigenomic cleavage sites. Interestingly, in contrast to all complete HDV genomes sequenced so far, the pt62 sequence was the only one with 3 bp mutations plus an insertion inside the 320 primer region (believed previously to be conserved) that was chosen initially to amplify R1 and R2 region sequences (Fig. 1). High variable regions comprise the non-coding part of the genome, the 3’-terminal part of the HD gene and, to a lesser extent, the 5’-terminal part of the HD gene. Secondary structure prediction using the mfold program yielded unbranched linear rod structures for pt26 and pt62 RNAs. The extremities of the rod corresponded to positions 800 and 1646. The secondary structure of the antigenomic region corresponding to the 3’ end of the HD gene (coding for the LHD) indicates that the editing site (amber/W) was conserved for all isolates (Fig. 4).
Coexistence of HDV I and II in Yakutia

Fig. 5. Deduced amino acid sequence of HDAg for two Yakutian isolates and reference sequences (see Fig. 3). Sequences were aligned using the CLUSTAL W (1.8) program. Sequences are numbered from the first amino acid residue. Residues identical to the pt26 sequence are marked with dots at aligned positions. Gaps were included to indicate missing amino acids. The different functional domains are indicated: the dimerization domain is simply underlined, the nuclear localization signal is marked with a thick line and the RNA-binding domain is doubly underlined. The LHDAg-packaging signal is indicated by an open box. The asterisk indicates the stop codon for sHDAg.

However, if base pairing surrounding this position was particularly strong for type I sequences (4 bp each side), as described previously (Casey et al., 1992; Polson et al., 1996), a less strong secondary structure was found within type II sequences (Fig. 4). Indeed, for all configurations obtained, a loop might have weakened the editing site, as only two C–G matches were observed between the editing site and the beginning of a 3 nt loop (Fig. 4).

The predicted LHDAg amino acid sequence of two Yakutian isolates was aligned with reference sequences (Fig. 5). On the amino acid level, pt62 and pt26 Yakutian HD proteins showed 89.8% strict identity with each other. Compared to reference HD amino acid sequences, pt62 and pt26 Yakutian HD proteins showed 83.7/86.6, 74.4/80.0, 74.1/74.5 and 63.4/63.4% sequence identity to types IIA, IIB, I and III, respectively. Most functional parts of the sHD protein, i.e. dimerization domain, nuclear localization signal and arginine-rich RNA-binding domain, were well conserved between all of the viruses analysed (Fig. 5). The most variable domains include the N-terminal part of HDAg and the C-terminal part of LHDAg. Inside this proline-rich carboxy extension, the cysteine residue C²¹¹ involved in the farnesylation anchoring was one of the three perfectly conserved amino acids among all HDAg protein sequences deduced; the two other amino acids were P²⁰¹ and Q²¹¹. However, the nuclear export signal

![Fig. 6. Phylogenetic analyses obtained from (a) ML analyses of sHD protein-coding region (phylogram) and (b) complete sequence MP bootstrap analysis (cladogram) from 19 HDV isolates, including the two Yakutian isolates. BV are indicated for (a) 10⁷ (ML) and (b) 10⁸ (MP) replicates. Scale represents the number of nucleotide substitutions per site.](https://example.com)
198LFPADPPFSQST described recently (Lee et al., 2000) was seen only for type I isolates, whereas the corresponding region of the Yakutian pt26 LHDAg was 198NPVPQGQL-PLLE210.

HDV nucleotide sequences encoding sHD as a marker of evolutionary genetics

Since sHDAg is essential for virus RNA replication and is type-specific, both the viral genome and the sHD gene might have evolved together with a specific link. We focused, therefore, on re-analysing the sHD protein-coding sequences to appraise the phylogenetic relationships of the newly characterized Yakutian virus sequences. By analysing 19 sHD protein-coding sequences through ML (Fig. 6a), MP and NJ analyses (data not shown), we confirmed that the monophyly of the Yakutian lineage, together with Taiwan3 and Japan, is robust (ML/MP/NJ BV = 98/99-55/96-6). In contrast, in such analyses, the Taiwan-TW2b sHD sequence is no more affiliated with the HDV type II clade. Instead, a common branching is observed for both the Taiwan-TW2b and the Peru1 sHD genes. The grouping of Taiwan-TW2b and Peru1 (while observed most frequently) occurs in only 49% of 100 ML replicates (Fig. 6a). This tree topology is also observed under MP and NJ analyses (10⁷/10⁵ replicates, MP/NJ BV = 75-60/54-5). All of these results may help to interpret cautiously the small MP BV obtained for R1 region analysis (10-5/10-6). HDV clade II comprises Japan, Taiwan3 and Yakutian sequences. Taiwan-TW2b and Peru1 group together, with MP/NJ BV under 10⁴ replicates of 63-36/51-94 %, respectively.

Discussion

Yakutia (Northern Siberia, Russia) is a region with a high prevalence of B and delta hepatitis. Indeed, the prevalence of hepatitis D is 3–5 times higher in Yakutia than in other regions of Russia. Therefore, it is important to investigate the genetic variability of HDV circulating in this endemic area.

Since 1993, it has been known that three major HDV genotypes exist (Wang et al., 1986; Imazeki et al., 1991; Casey et al., 1993). Because of the absence of trans-complementation between HDV types I and III, Casey & Gerin (1998) argue that these viruses are not only of a different genotype but also of a different ‘type’. Using RFLP and genome sequencing associated with phylogenetic analyses, we demonstrated that at least two HDV genotypes, I and II, are currently present in Yakutia. It was predictable to find genotype I in Yakutia, because all Russian isolates described so far (Ryzhova et al., 1997; Flodgren et al., 2000) are affiliated with this genotype. In contrast, the discovery of genotype II in Yakutia was both unexpected and surprising, since the novel restriction profile observed among 15 isolates (Table 1) might be more likely derived from type I isolates (acquisition of one SmaI site) than from type II isolates (acquisition of two SmaI sites). HDV genotype II is described classically as being restricted to Far East Asia and is found only in Taiwan and Japan (Imazeki et al., 1990; Lee et al., 1996; Wu et al., 1998; Sakugawa et al., 1999). Our findings indicate that HDV genotype II should be considered to be more widespread. Unlike in East Asia, where type II prevails over type I (Wu et al., 1995), an equal proportion of both genotypes was observed among hospitalized patients in Yakutia. Interestingly, in this area, type II isolates are less distantly related to each other than type I isolates (Fig. 2). This observation may suggest that different type I isolates might have been introduced repeatedly into this area during the last century when the migration of individuals from the European part of Russia to Yakutia increased significantly. Another possible explanation is that type I isolates have been present in this area for a very long time and that type II isolates were introduced recently. The first hypothesis is probably supported by the fact that all of the Russian patients in our study (4 out of 29) were infected with type I viruses. The coexistence of two genotypes in this area might represent a ground for HDV inter-type recombination; however, we could detect in this preliminary study neither mixed infections (Wu et al., 1999) nor virus recombination between type I and II viruses.

The pathogenesis of HDV liver disease remains intriguing and complex. It might involve specific HDV features (expression and replication, genetic variability, ribozymes, etc.) and specific HBV features (genotypes, mutants, level of replication, etc.) in the context of host genetic background, immune response and superimposed cofactors (such as alcohol and other viruses). Wu et al. (1995) and Sakugawa et al. (1999) discovered that in Taiwan or in the Miyako islands, HDV genotype II seemed to be associated with a lower level of pathogenicity than HDV genotype I. In hospitalized patients from Yakutia, we observed that both genotype I and genotype II infections could lead to severe infections.

Other factors could account for the differences in the different cohorts of patients studied. First, in Taiwan, the difference of pathogenesis between HDV types I and II is significant for fulminant hepatitis associated mainly with HDV type I. Interestingly, a comparison of the secondary structures obtained suggests that base pairing surrounding the editing site seems to be less strong for type II than for type I sequences (Fig. 4). Studies from type I sequences indicate that conservation of these complementary base pairs is crucial for editing efficiency (Casey et al., 1992; Polson et al., 1996). HDV fulminant hepatitis might be associated with the efficiency of HDV type I editing, leading to swift virus dissemination during acute hepatitis. The characterization of Yakutian HDV viruses linked to HDV fulminant hepatitis and in vitro studies might help to test such a hypothesis. Second, the distribution of HDV genotypes may be connected to the distribution of HBV genotypes. In the south eastern part of European Russia, a high incidence of acute and fulminant HBV genotype D has
been observed recently and is associated with infection with HDV genotype I in 29% of cases (Flodgren et al., 2000). Whether or not a specific link exists for Yakutian HDV type II isolates with HBV genotype B or C isolates that are predominant in Asia (Theamboonlers et al., 1999; Orito et al., 2001) requires further study.

The original population of Yakutia is believed to have been present there for the past 40,000 years. The characterization of HDV lineages might be a useful tool to follow population migration, as it has been suggested for HBV, human T-lymphotropic virus type I and JC viruses (Agostini et al., 1997; Blitz et al., 1998; Van Dooren et al., 1998). HDV phylogenetic studies have been based almost exclusively on the analysis of the 3′-terminal part of the HD gene (Fig. 2a). The low similarity of this region might lead to erroneous HDV lineage reconstruction. Indeed, while being functional for assembly, LHDAg is not necessary for RNA replication, but inhibits it. Recombinations (Wu et al., 1999a) or deletions (P. Dény, unpublished) of parts of the genome have also been characterized. In contrast to LHD, the product of the sHD gene is essential for RNA replication and seems also to be type-specific (at least for studies involving complementation between types I and III). We suggest, therefore, that the coding sequence for sHD is a more appropriate marker for evolutionary genetics than the region encoding the carboxy-terminal extension. By characterizing new HDV isolates from Yakutia and using extensive phylogenetic approaches on various genomic regions, we demonstrated that all type II Yakutian sequences might be derived from a common ancestor and that all of these viruses had probably evolved from a common ‘Asian and Siberian’ HDV type II prototype. Although our results based on the sHD gene and complete genome phylogenetic studies are preliminary (Fig. 6), they could also suggest that Taiwan-TW2b (Wu et al., 1998) (proposed to be the prototype of HDV subtype IIb) might no longer be affiliated with HDV clade II. This suggestion reflects a wider HDV variability than thought previously.

We thank J. C. Trinchet, M. Milinkovitch, P. Anaïs and E. Gordien for helpful discussions. We are particularly grateful to J. Casey for giving his personal comments and suggestions on an earlier version of the manuscript. V. I. was supported by a ‘bonus qualité recherche’ grant from Université Paris 13. N.R. is partly supported by the ‘Association Nationale de Recherches sur le Sida’.

References


Received 2 April 2001; Accepted 23 July 2001