Complete genome analysis of Delta hepatitis from Afghanistan.

Abbas Ali Husseini

1Istanbul Gelisim University

May 16, 2023

Abstract

There is a lack of information on the complete genomic profile of the hepatitis delta virus in Afghanistan. Therefore, the main objective of this study is to determine the molecular profile of complete genome of hepatitis delta virus in individuals who are positive for the hepatitis B surface antigen in the general population of Afghanistan. A total of 234 hepatitis B antigen-positive sera were analyzed by chemiluminescence microparticle immunoassay to detect anti-delta antibodies. Serologically positive samples were later confirmed by real-time polymerase chain reaction assay. Subsequently, the whole genome was amplified using two set of primers generate two partially overlapping segments and later sequenced. Phylogenetic comparison and evolutionary divergence analysis was performed with the MEGA7 software package. The coinfection rate among hepatitis B antigen carriers in Afghanistan was 2.1%. Finally, the whole genome of 4 hepatitis delta virus isolates from Afghanistan was successfully amplified, all of which were later assigned to genotype I in phylogenetic analysis. The amino acid composition of the hepatitis delta antigen and its functional motifs were consistent with genotype 1, but the subtype-specific amino acid signature illustrated different potential subtypes.

Introduction

Hepatitis delta virus (HDV) is a defective ribonucleic acid (RNA) virus that requires the presence of hepatitis B virus (HBV) to replicate and cause disease. In patients with chronic HBV infection, HDV infection may develop as a superinfection or coinfection with HBV. The worldwide prevalence of HDV infection is not known because many countries do not routinely test for HDV infection. However, it is estimated that approximately 5% to 20% of people chronically infected with HBV are also infected with HDV, depending on the geographic region. Moldova, Mongolia, and countries in West and Central Africa are among the geographic hotspots identified by the World Health Organization as having the highest prevalence of HDV infection. In these regions, HDV infection may be responsible for a substantial proportion of cases of liver disease and liver-related mortality.

HDV infection can lead to more severe liver disease than HBV infection alone and is also associated with HBV complications. Chronic HDV infection is associated with a higher risk of cirrhosis, liver failure, and liver cancer and may lead to more rapid progression of liver disease. Depending on the host and viral circumstances, HBV/HDV coinfection can be a serious disease with severe consequences. In addition to the presence of other liver diseases or risk factors such as alcohol consumption, obesity, or diabetes, and the age of the individual at the time of infection, the diversity of viral nucleotide sequences is one of the main variables affecting disease development and manifestation. The eight HDV genotypes (I- VIII) with global ethno-geographic distribution are characterized by viral nucleotide sequence variations of more than 19% to 38% throughout the HDV genome.

Numerous studies have investigated how genotypes affect disease progression. Different HBV genotypes are associated with a range of clinical symptoms, including the severity and longevity of liver disease, the incidence of hepatocellular carcinoma, response to interferon therapy, and even the mode of transmission and liver transplantation. The clinical impact of HDV-1 can range from mild to severe disease. While HDV-3
can cause fulminant hepatitis, HDV-2 is typically associated with a milder form of the disease. There are limited data on the clinical course of the other five HDV genotypes.

One of the most important public health problems in low-income countries such as Afghanistan is this problem caused by HBV/HDV coinfection. Studies of viral hepatitis, particularly HDV, are incredibly rare in this country. Our prior studies suggest that HDV infection is an important public health problem in Afghanistan, particularly among people with chronic hepatitis B, people who inject drugs, and pregnant women. Sequence analyzes undoubtedly contribute significantly to our knowledge of the molecular profile of the virus of interest. Studies on seroprevalence and molecular characterization of infectious diseases are crucial for the development of appropriate and successful public health interventions. Undoubtedly, the molecular epidemiology of HDV in Afghanistan remains to be elucidated. Due to the lack of such information in Afghanistan, the main objective of the current study is to assess the current molecular profile of the complete HDV genome among hepatitis B antigen (HBsAg) carriers in this country. In particular, we are concerned with the genotypic distribution and divergence analysis of sequences, especially functional motifs in the large hepatitis delta antigen (L-HDAg).

Material and methods

Site of the study and sampling procedures

To test for HBV seromarkers using point-of-care rapid tests (Standard Diagnostics, Korea/USA), the National Public Health Institute of Afghanistan and the World Health Organization randomly collected 5897 whole blood samples from adults aged 25-70 years from Nangarhar (eastern zone), Herat (western zone), Mazar-e-Sharif (northern zone), Kandahar (southern zone), and Kabul (central zone) (A. Husseini et al., 2019). HBsAg-positive whole blood samples were collected in tubes containing K3EDTA. Serum was separated from whole blood by centrifugation by spinning the tubes at 1200 rpm for 10 minutes. Serum was separated after centrifugation and stored at -18°C. The study included 234 HBsAg-positive individual sera (143 males and 91 females, mean age 40.4±14.5 years) identified from 5897 randomly selected samples and subsequently analyzed using a chemiluminescent microparticle immunoassay (Abbott Laboratories, Illinois, USA). Tubes were spun at 1200 rpm for 10 minutes to separate serum from whole blood during centrifugation. After centrifugation, the isolated serum was stored at -18°C.

Serological anti-HDV tests

All 234 HBsAg-positive samples were subjected to anti-HDV antibody screening using a chemiluminescent microparticle immunoassay (Abbott Laboratories, Illinois, USA) according to the manufacturer’s instructions.

Viral nucleic acid extraction and viral load measurement

Viral nucleic acids were extracted from 150 μl of patient sera using a spin column-based purification procedure via a commercial AUGEN˙IX nucleic acid isolation kit (AUGEN˙IX, Turkey) according to the manufacturer’s instructions. Subsequently, HDV RNA was quantified in real time using a previously published in-house HDV method based on polymerase chain reaction (PCR).

PCR amplification and genome sequencing

Two primer sets, previously designed to generate two partially overlapping segments of 950 and 1100 base pairs covering the entire HDV genome, were used to amplify the complete circular HDV RNA genome.

Complementary DNA (cDNA) of each segment was synthesized using a reaction mixture consisting of 5 μl of eluted RNA, 5 μl of 2 mM dNTPmix (Fermentas, Vilnius, Lithuania), 10 pmol of reverse primer in 2.5 μl, 4 μl of 5X reaction buffer, 4U AMV RT-PCR (Roche Diagnostics, Germany) and 3.5 μl of ddH2O in a total volume of 20 μl. Then, the reaction mixture was incubated at 95°C for 3 minutes and at 42°C for 1 hour to complete the cDNA synthesis. Next, 5 μl of cDNA was mixed with 10 pmol of each primer (2.5 μl each forward and reverse), 5 μl of 2 mM dNTPmix, 2 μl of 25 mM MgCl2, 2.5 U of Taq DNA polymerase (Roche
Diagnostics, Germany), 19 μl of buffer, and ddH2O for a total volume of 50 μl. Thermocycling conditions were previously optimized and applied in this study as before.

Amplicons were visualized using 1% agarose gel electrophoresis stained with ethidium bromide. PCR reactions were performed with appropriate negative controls to avoid false positive results. To avoid false-positive results, PCR reactions were performed with the appropriate negative controls. Before sequencing, PCR fragments were purified using a silica-based technique, and all primer sets were used for commercial sequencing.

**Nucleic acid and amino acid sequence analysis**

All sequence isolates obtained from HDV-positive sera were manually processed, aligned, and compared with 25 reference sequences from the GenBank database representing eight widely used and known genotypes worldwide.

Accession numbers of reference sequences are as follows: Genotype I (X85253, U81989, U81988, M58629, D01075, HQ005372), genotype II (U19598, AJ309880, AF104264), genotype III (AB037948, AB037947), genotype IV (AF309420, AF209859, AB118847, AB088679), genotype V (AJ584848, AM183331, AM183326), genotype VI (AM183329, AM183332), genotype VII (AM183333, AX741169), genotype VIII (AM183330, AM183327, AX741169) respectively.

Phylogenetic comparison was performed using distance matrix/UPGMA, maximum likelihood, and neighbor-joining methods with bootstrap resampling (1000 replicates) to confirm the reliability of the phylogenetic tree using Kimura 2 parameters through the MEGA7 software package (Kumar et al. Institute for Genomics and Evolutionary Medicine, Temple University). In addition, evolutionary divergence analysis of the isolates was performed using the maximum composite likelihood model. The nucleotide and L-HDAg peptide amino acid compositions of the whole genome were also computed.

The amino acid sequences of the HDAg from Afghan isolates were compared to reference sequences from all genotypes. The accession numbers of the reference sequences are as follows: Genotype I (QGQ76699, AXF50994, QGQ76701, AJ072430, AJ072440, CAQ16911, CAQ16913), genotype II (QDC33395, AXT99879, SCC98302), genotype III (AIR77045, AIR77037), genotype IV (ABO87258, ABO87259), genotype V (CAE51160, CAE51163, CAJ66090), genotype VI (SCC98303, CAJ66093), genotype VII (AXF51015, AXF50952), genotype VIII (SCC98324, SCC98323) respectively.

The tertiary structure of the HDAg peptide from Afghan isolates was determined using I-TASSER (Iterative Threading ASSEmbly Refinement), a protein structure prediction and structure-based function annotation tool.

**Results**

Chemiluminescence microparticle immunoassay detected coinfection with HDV in five individuals with a mean age of 49.25 years from 234 HBsAg-positive samples. All anti-HDV positive samples were confirmed by real-time PCR, with viral loads ranging from 5.3x103 to 2.2x104 IU/ml. In four of the samples, the entire hepatitis delta genome was actually amplified. In subsequent phylogenetic analysis, all belonged to the genotype I branch in trees constructed using distance matrix/UPGMA, maximum likelihood, and neighbour-joining modules (Figure 1, 2, and 3) and stored at GenBank under accession numbers OQ863057, OQ863058, OQ863059, and OQ863060, respectively. The complete HDV genome sequence of the Afghan isolates has a length of 1694 nucleotides with a GC content of 58.1-59.5% (Table 1). Based on the intragenomic evolutionary divergence analysis, the sequence similarity between the four Afghan isolates ranged from 88% to 92.7% (Table 2). In addition, the overall genomic homogeneity of the Afghan isolates was 89%. However, the similarity of large and small HDAg was 84% and 82%, respectively. On the other hand, the mean intergenotypic nucleotide distance between the Afghan HDV isolates and the reference HDV genotype I isolates calculated using the Kimura two-parameter algorithm in the MEGA software was 15%. The sequence most closely related to the Afghan isolates was the D01075 genotype 1 sequence (19.5% divergence), whereas the AB037947 genotype III sequence was the most distant from the Afghan isolate OQ863060 (48.5% divergence) (supplemental document 1). The predicted HDAg amino acid sequence of the Afghan isolates was aligned
with reference sequences. At the amino acid level, the intergenotypic divergence is generally less than 13% between the Afghan isolates, with the exception of OQ863057, where the overall divergence was over 25.5% compared to other Afghan isolates. The highest similarity was between isolates OQ863059 and OQ863060 (90.2%) (Supplementary Document 2). The genotype III has the highest distance in HDAg amino acid sequences compared to all other genotypes. The amino acid sequences of large HDAg from each isolate were aligned and motifs associated with functional properties were examined. The RNA-binding domain (RBD) was located at the N-terminus of the protein sequences. The RBD appears to be highly conservative in all isolates except OQ863057. The RBD is rich in arginine and glutamic acid. The coiled-coil domain is rich in lysine and leucine, which are located at the N-terminus of the sequence, and the residues are highly conserved between genotypes. The nuclear localization signal showed the least divergence. However, the most diverged domain was the helix-loop-helix, which is rich in glutamic acid, lysine, and leucine. A virus assembly signal present in the carboxy-terminal region of L-HDAg was highly conserved among the Afghan isolates (Figure 4). Genotype comparison of L-HDAg shows that the N-terminal RBD varies greatly between genotypes. In contrast, the nuclear localization signal (NLS) appears to be highly conserved among genotypes. The packaging signal is highly conserved in genotype 1 but differs greatly compared to the other genotypes. L-HDAg is rich in arginine, glutamic acid, glycine, proline and lysine, which account for over 50% of the sequence. Approximately 45% of all amino acids that make up HDAg are hydrophobic. The amino acid composition of phenylalanine and lysine is relatively high in genotype 1 compared to the other genotypes. On the other hand, the ratio of amino acid to proline composition appears to be a genotype-specific variation (Figure 5). The glycine residues are very well conserved between genotypes and especially between genotype 1. The predicted tertiary structure of the L-HDAg amino acid revealed a dimer structure.

Discussion

This is the first meticulous analysis of the complete HDV genome from Afghan isolates aimed at determining the molecular profile of HDV prevalent in the Afghan population. As the previous partial HDAg analysis of the same isolates revealed, the phylogenetic analysis of the complete genome in the current study also determines that HDV genotype 1 is the most common variant circulating in the Afghan population, with a relative frequency of 0.08% in the general population and 2.1% in HBsAg carriers. The estimated anti-HDV prevalence among HBsAg-positive individuals worldwide is 4.5% (95% CI 3.6-5.7), ranging from 26% in central Africa to 0.05% in eastern and southern Africa.

Afghanistan is a landlocked country located in South and Central Asia. Pakistan forms the eastern and southern borders, Iran the western, Turkmenistan, Uzbekistan, and Tajikistan the northern, and China the northeastern. Of 234 HBsAg-positive samples, chemiluminescence microparticle immunoassay detected coinfection with HDV in five individuals with an average age of 49.25 years. All anti-HDV positive samples with viral loads ranging from 5.3 x 103 to 2.2 x 104 IU/ml were confirmed by real-time PCR. In four of the samples, the entire hepatitis delta genome was amplified. In this terrestrial area, the prevalence of HDV appears to vary by region. The prevalence of HDV in China is still unclear and varies by ecoregion. While one study found a nationwide HDV seroprevalence between 1% and 2%, other multicenter studies have found that HDV is significantly more prevalent in autonomous regions. Similarly, HDV prevalence in Iran shows a range from 3.9% to 17%. Studies from Pakistan found Pakistan to be a highly endemic region with an overall prevalence of approximately 16.6%, ranging from 8.5% to 88%. Little information is available from neighboring countries to the north. In Central Asia, in the northern part of Afghanistan, the seroprevalence of HDV infection was 8.3% in the general population and 51.3% in patients who tested positive for HBsAg. A study from Uzbekistan reported a high burden of HDV in that country (49.1% in HBV-positive patients). Similarly, a study showed HBV/HDV co-infection of 11.2% among Tajik HBV-positive patients. However, in the current study, the anti-HDV prevalence among Afghan HBsAg-positive patients appears to be lower than in neighboring countries in this region. This conclusion may be deceptive because of the overall sample size and target population. HDV is subdivided into genotypes HDV-1 to HDV-8, whose distribution is geographically limited except for the globally predominant HDV-1.

Among others, genotypes 2 and 4 are mainly found in Asia. In Afghanistan, all HDV patients are infected
with genotype 1, which is associated with the greatest pathogenic impact in HBV/HDV-coinfected patients, and the genotype distribution appears to be consistent with that in neighboring countries.

The complete HDV genome sequence of the Afghan isolates was 1694 nucleotides long and had a GC content of 58.1%-59.5%. Several studies suggest that the GC content of a genome may be correlated with various biological traits such as growth temperature in bacteria genome size and host genomic GC. In the current analysis, a negative correlation was found between GC content and genome size (r: -0.5 P value: 0.01). However, there is no empirical evidence that the GC content of HDV viral genomes plausibly affects the biological properties of the virus, such as the ability to replicate and infect host cells. In addition, pairwise distance calculation showed high similarity between Afghan isolates OQ863058, OQ863059, and OQ863060 (91.6%), which may indicate a common evolutionary origin. However, isolate OQ863057 had 86% similarity, likely indicating a different origin. The HDV subtype-specific amino acid signature suggests that different subtypes may circulate in the Afghan population.

HDAg is a protein encoded by the HDV genome. HDV replication occurs via a dual rolling circle mechanism involving the synthesis of antigenomic RNA intermediates. HDAg is essential for HDV replication and virus assembly. It is encoded as two isoforms in a segment of the HDV genome, namely small HDAg (S-HDAg, 195 aa, 24 kDa) and large HDAg (L-HDAg, 214 aa, 27 kDa). While S-HDAg is required for initiation and maintenance of replication, L-HDAg negatively regulates replication and triggers envelopment of the virus in HBV surface proteins. Several functional motifs are at work in HDAg, including RNA-binding domains, a coiled-coil domain, a nuclear localization signal (NLS), a helix-loop-helix, and a packaging signal domain, which are important for survival and regulation of various aspects of the viral life cycle. It is likely that variations in the functional motifs of the HDAg protein may influence their functions. The RNA-binding domains of HDAg are responsible for binding to the HDV RNA genome and the antigenome, which in turn are involved in several steps in the viral life cycle.

The coiled-coil and helix-loop-helix (HLH) domains are structural motifs present in many proteins. The coiled-coil domain of HDAg is involved in the formation of dimers and oligomers through the association of two or more alpha helices that wrap around each other in a supercoil, directing and facilitating protein-protein interactions. The coiled-coil domain of HDAg is also involved in interaction with the hepatitis delta virus (HDV) RNA genome. The stability of coiled-coil structures is largely determined by the interactions between amino acid side chains located at the helix interface. The coiled-coil amino acid residues of HDAg are predominantly formed by hydrophobic and charged amino acids. In particular, the hydrophobic cores of the coiled coils in HDAg tend to be built from aliphatic amino acids (L, I, V, and A) in all genotypes. Leucine residues were very well conserved in the genomics, suggesting that these residues are very important for the formation of coiled coils in HDAg. In addition to hydrophobic residues, there is a preference for positively (Lys, Arg) and negatively (Glu) charged residues in HDAg coiled-coil domains, which in turn may contribute to the stability of coiled-coil structures by forming electrostatic interactions or salt bridges. Lysine is a positively charged amino acid that can form salt bridges with negatively charged amino acids such as glutamic acid.

Leucine is a hydrophobic amino acid that is often found in the hydrophobic core of coiled-coil domains. Therefore, mutations in such residues may affect coiled-coil stability. In the isolates studied, Glu was replaced by aspartic acid in one isolate. Lys was also replaced by arginine and threonine. Tertiary structure prediction analysis of the sequences showed that the domain is relatively stable. Glu and Asp, both amino acids, are acidic, that is, they have a negatively charged carboxyl group (-COOH) at one end of the molecule. Lysine and arginine are positively charged amino acids, while threonine is neutral. Therefore, the divergence in the Afghan isolates does not appear to have any functional implications.

The NLS domain in HDAg is responsible for the transport of HDAg from the cytoplasm to the nucleus. The NLS domain is located in the N-terminal third of HDAg. Variations within the nuclear localization signal (NLS) domain in HDAg may affect its function. One study reported that different nuclear localization signals have different effects on protein function. For example, variations in the NLS domain dramatically suppress HIV-1 gene expression and virus production. The NLS domain of Afghan isolates is rich in proline and
arginine, and the relevant residues are highly conserved not only in genotype 1 but in all known genotypes.

The packaging signal domain is a conserved RNA sequence required for encapsulation of the HDV genome in virions. Despite the retention of highly conserved functional motifs, it is not known whether sequence divergence between and within genotypes can affect biological properties such as replication efficiency of different HDV genotypes.

The amino acid composition of HDAg from Afghan isolates was studied and found to contain a high proportion of hydrophobic amino acids (approximately 45%). Hydrophobic amino acids play an important role in the structure and function of peptides. They are usually found inside proteins and peptides, where they help stabilize the structure by forming hydrophobic interactions with other hydrophobic amino acids. The hydrophobic nature of HDAg is thought to be important for its function in HDV replication. In addition, L-HDAg is rich in charged amino acids (approximately 44%), which in turn are important contributors to HDAg stability and its function.

The high conservation of glycine in HDAg is probably due to its structural and functional importance in protein. Glycine is the smallest amino acid that is very flexible and can adopt a wide range of conformations due to its small size and lack of a bulky side chain. In addition, glycine is frequently found in turns and loop regions of proteins. This flexibility and structural role is most likely important during the life cycle of HDV. Conclusion: Empirical studies have shown that the most common HDV genotype circulating in Afghanistan is genotype 1, which may be heterogeneous in terms of subtype. However, this study paves the way to understanding the molecular dynamics of hepatitis Delta in this terrestrial area and may contribute to disease control by providing information on viral genetic diversity and its evolutionary history. It can also help track the spread of the disease and identify outbreaks.

Ethics Committee Approval: Ethics committee approval was received for this study from the Afghanistan National Public Health Institute, Feb/10/2016, 361537.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: This study was supported by Istanbul Gelisim University, the office of Scientific Research Projects, Grant no: Grant No. KAP-050421-AHH.

References
Figure 1) Phylogenetic tree obtained by distance matrix/UPGMA comparison (with Kimura-2-correction) after bootstrapping 1000 replicates of HDV genomic RNA sequence.
Figure 2) Phylogenetic tree obtained by Maximum likelihood method (with Kimura-2correction) after bootstrapping 1000 replicates of HDV genomic RNA sequence.
Figure 3) Phylogenetic tree obtained by Neighbor-Joining method (with Kimura-2 correction) after bootstrapping 1000 replicates of HDV genomic RNA sequence.

Figure 4) Functional domains of predicted L-HDAg protein sequences
Figure 5) Amino acid composition of L-HDAg among different HDV genotypes

![Figure 5](image7.emf)

Figure 6) The predicted tertiary structure of L-HDAg amino acid sequences of Afghan isolates

Table 1) Nucleotide composition of Hepatitis Delta virus complete genome of Afghan isolates

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>OQ863057</th>
<th>OQ863058</th>
<th>OQ863059</th>
<th>OQ863060</th>
</tr>
</thead>
<tbody>
<tr>
<td>T (U) %</td>
<td>19.7</td>
<td>19.8</td>
<td>21.0</td>
<td>20.3</td>
</tr>
<tr>
<td>C %</td>
<td>30.3</td>
<td>29.5</td>
<td>30.5</td>
<td>30.3</td>
</tr>
<tr>
<td>A %</td>
<td>20.8</td>
<td>20.9</td>
<td>20.7</td>
<td>20.5</td>
</tr>
<tr>
<td>G %</td>
<td>29.2</td>
<td>29.9</td>
<td>28.6</td>
<td>29.5</td>
</tr>
<tr>
<td>GC %</td>
<td>59.5</td>
<td>59.4</td>
<td>58.1</td>
<td>59.7</td>
</tr>
<tr>
<td>Total</td>
<td>1694</td>
<td>1693</td>
<td>1693</td>
<td>1694</td>
</tr>
</tbody>
</table>

Table 2) Estimates of evolutionary divergence of Afghan isolates and comparison with genotype 1 reference sequences

<table>
<thead>
<tr>
<th>Genotype</th>
<th>OQ863057</th>
<th>OQ863058</th>
<th>OQ863059</th>
<th>OQ863060</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>29.3</td>
<td>29.6</td>
<td>29.3</td>
<td>29.3</td>
</tr>
</tbody>
</table>

Hosted file