ORIGINAL ARTICLE VIROLOGY

Comprehensive analysis of mutations in the hepatitis delta virus genome based on full-length sequencing in a nationwide cohort study and evolutionary pattern during disease progression

E. Shirvani-Dastgerdi[1], S. Amini-Bavil-Olyaee[2], S. Moayed Alavian[3], C. Trautwein[1] and F. Tacke[1]

1) *Department of Medicine III, RWTH-University Hospital Aachen, Aachen, Germany,* 2) *Department of Molecular Microbiology and Immunology, Keck School of Medicine, University of Southern California, Harlyne J. Norris Cancer Research Tower, Los Angeles, CA, USA and* 3) *Baqiyatallah Research Centre for Gastroenterology and Liver Diseases, Baqiyatallah University of Medical Sciences, Tehran, Iran*

Abstract

Delta hepatitis, caused by co-infection or super-infection of hepatitis D virus (HDV) in hepatitis B virus (HBV) -infected patients, is the most severe form of chronic hepatitis, often progressing to liver cirrhosis and liver failure. Although 15 million individuals are affected worldwide, molecular data on the HDV genome and its proteins, small and large delta antigen (S-/L-HDAg), are limited. We therefore conducted a nationwide study in HBV-HDV-infected patients from Iran and successfully amplified 38 HDV full genomes and 44 L-HDAg sequences from 34 individuals. Phylogenetic analyses of full-length HDV and L-HDAg isolates revealed that all strains clustered with genotype 1 and showed high genotypic distances to HDV genotypes 2 to 8, with a maximal distance to genotype 3. Longitudinal analyses in individual patients indicated a reverse evolutionary trend, especially in L-HDAg amino acid composition, over time. Besides multiple sequence variations in the hypervariable region of HDV, nucleotide substitutions preferentially occurred in the stabilizing P4 domain of the HDV ribozyme. A high rate of single amino acid changes was detected in structural parts of L-HDAg, whereas its post-translational modification sites were highly conserved. Interestingly, several non-synonymous mutations were positively selected that affected immunogenic epitopes of L-HDAg towards CD8 T-cell- and B-cell-driven immune responses. Hence, our comprehensive molecular analysis comprising a nationwide cohort revealed phylogenetic relationships and provided insight into viral evolution within individual hosts. Moreover, preferential areas of frequent mutations in the HDV ribozyme and antigen protein were determined in this study.

Keywords: Delta virus, hepatitis B virus, hepatitis D antigen, hepatitis, phylogenetic tree, ribozyme, sequencing, viral evolution
Original Submission: 15 September 2014; **Revised Submission:** 21 November 2014; **Accepted:** 18 December 2014
Editor: G. Antonelli
Article published online: XXX

Corresponding author: F. Tacke, Department of Medicine III, RWTH-University Hospital Aachen, Aachen, Germany
E-mail: frank.tacke@gmx.net

Introduction

Hepatitis delta virus (HDV) is a defective virus with a covalently closed circular single-stranded RNA genome that is the only member of *Deltavirus* genus [1]. HDV infection is strictly associated with hepatitis B virus (HBV) to obtain its envelope from hepatitis B surface proteins (HBsAg). The viral genome size is about 1.7 kb, which encodes two proteins designated as HDV small D antigen (S-HDAg) and HDV large D antigen (L-HDAg) [1–3]. Once HDV enters the cell, its genomic RNA is transcribed into intermediate anti-genomic RNA by cellular transcription enzymes [2,3]. Both genomic and anti-genomic strands have three different regions, including hypervariable, autocatalytic and HDAg coding domains. Considering non-conserved sequences of the hypervariable region and the well-preserved structure of the catalytic domain [4], investigators agreed to define HDV genotypes by comparing HDAg coding domains. However, the best evolutionary

http://dx.doi.org/10.1016/j.cmi.2014.12.008

information (reference standard) for HDV is obtained by full-length genome sequence analyses [5–7]. Phylogenetically, eight different clades of HDV have been introduced (genotypes 1 to 8) from different parts of the world, with distinct geographical regions and different outcome of diseases [8].

Globally, more than 15 million people have been co-infected with HBV and HDV [9]. Due to global HBV vaccination programmes since the 1990s, HDV infection has been relatively well controlled in Europe, especially Western Europe [1]. In contrast, several countries in the Middle East are considered endemic regions for HDV infections [9–13]. Despite partial or full genome analyses of HDV reported from distinct geographical areas [14,15], there are limited data available on the molecular evolution of HDV from this region, particularly Iran [9,16–19]. In the current cross-sectional and longitudinal study, we analysed the full-length genome as well as the entire HDAg coding sequence of HDV isolates from a large set of HDV antibody positive patients (n = 34) from different parts of Iran. These analyses aimed to investigate the molecular epidemiology of HDV full-length genome sequences and their evolutionary pattern in a large cohort.

Materials and methods

Patients

Seventy-one serum samples from 49 Iranian patients with chronic HBV-HDV co-infection, who were referred to Tehran Hepatitis Centre, were included in this study. HBsAg-positive and HBeAg-negative patients with positive IgG for HDAg were consecutively included. The Tehran Hepatitis Centre physician team prepared a questionnaire and recorded available risk factors, virological, clinical, histological and serological data. Written informed consent was obtained from all participants, and the ethics committees of the Baqiyatallah Research Centre for Gastroenterology and Liver Diseases (Iran) and the RWTH University Hospital Aachen (Germany) approved the study. All patients tested negative for hepatitis C virus (HCV) and human immunodeficiency virus (HIV) antibodies.

HDV genome sequencing

HDV RNA was extracted from sera by the QIAamp viral RNA mini kit (Qiagen, Hilden, Germany) and then subjected to cDNA synthesis using a Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany), followed by HDV whole genome amplification through a nested-PCR method as described previously [20]. Amplicons were purified using the QIAquick PCR purification kit (Qiagen) and subjected to bidirectional primer-walking sequencing using the 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Evolutionary and phylogenetic analysis

All sequenced fragments were assembled using the Contig Assembly Program of the BioEdit Sequence Alignment Editor, version 7.2.0 [21]. To investigate the mutational pattern of the HDV ribozyme domain and L-HDAg amino acids, the sequences were compared with each other and with the same domains of other genotypes of HDV [20]. To perform the evolutionary and phylogenetic analysis, multiple sequence alignments were carried out for the assembled full-length as well as L-HDAg encoding sequences together with defined HDV genotypes 1 to 8 as reference genes using the MEGA software, version 5.2 [22]. The Kimura two-parameter algorithm was used to compute evolutionary distances using MEGA software [23]. Phylogenetic relationships, and evolutionary rates were also calculated with the neighbour-joining method using 1000 replicate bootstrap value by MEGA software and confirmed by maximum likelihood analysis [23,24].

Selective pressure analysis

The numbers of non-synonymous and synonymous substitutions per site (dN and dS, respectively) for each codon were estimated to investigate the impact of selective pressure on the protein. All studied cases were aligned based on their nucleotide sequences and applied for selection analysis by VarPlot software (http://sray.med.som.jhmi.edu/SCRoftware/VarPlot/) as described elsewhere [25]. The Nei–Gojobori method was used for the dN/dS calculations. A dN/dS ratio of 1 is indicative of the same levels of synonymous and non-synonymous mutations at each position, which leads to the lack of selective pressure at these sites known as neutral evolution. The ratios of dN/dS >1 indicate a fitness advantage of non-synonymous changes and therefore a positive selection and for dN/dS <1 a negative selection is expected due to the detrimental effect of non-synonymous mutations at those sites [25,26].

Reference numbers of HDV sequences

HDV genotype reference sequences retrieved from GenBank were genotype 1: AY633627.1_Iran, U81989.1_Ethiopia, X04451.1_Italy, M84917.1_Lebanon, M58629.1_Nauru, U81988.1_Somali, M92448.1_Taiwan, D01075.1_US1, and L22066.1_US2; genotype 2: AB088679.1_Japan; genotype 3: L22063.1_Peru; genotype 4: AF018077.1_Taiwan; genotype 5: AX741149.1_dFr47; genotype 6: JA417561.1_dFr48; genotype 7: JA417541.1_dFr45; and genotype 8: JA417566.1_dFr644. Sequences of the studied isolates were submitted to GenBank and are available under Accession numbers KJ744214 to KJ744257.

Please cite this article in press as: Shirvani-Dastgerdi E, et al., Comprehensive analysis of mutations in the hepatitis delta virus genome based on full-length sequencing in a nationwide cohort study and evolutionary pattern during disease progression, Clinical Microbiology and Infection (2015), http://dx.doi.org/10.1016/j.cmi.2014.12.008