INT. J. BIOL. BIOTECH., 16 (3): 613-621, 2019.

GENOTYPING OF LOCAL ISOLATES OF HEPATITIS DELTA VIRUS, USING PHYLOGENETIC ANALYSIS

Narjis Naz[1], Shaimuna Fareeha Sajjad[2], Syed Ejaz Alam[2], Waquaruddin Ahmed[2], Huma Qureshi[2] and Obaid Yusuf Khan[1*]

[1]*Department of Genetics, University of Karachi, Karachi, Karachi-75270, Pakistan*
[2]*Pakistan Health Research Council (PHRC) Research Center, Jinnah Postgraduate Medical Centre, Karachi, Pakistan*
[*] Corresponding author email: oykhan@uok.edu.pk Tel.: +92-334-375-2368.

ABSTRACT

Hepatitis B (HBV) and Hepatitis D (HDV) is a confronting health issue in Pakistan with the number of individuals infected with HDV is not showing any decline. Prevalence of coinfection of HDV/HBV is about 16 – 57% in Pakistan. HDV being a satellite virus of HBV, replicates only in the presence of HBV. Compared to chronic carrier state of HBV alone, the HBV-HDV coinfection or superinfection are more susceptible to result in hepatocellular carcinoma (HCC), causing threefold increase in fatality. HDV has been classified into eight distinct clades or genotypes with clade I being predominant in Pakistan. Genotyping of viral infection is a routine diagnosis process as different clades have their own pattern and progression of disease hence information about the genotype of an infection helps in the prognosis of the disease. Conventionally, genotyping was done by RFLP analysis of the R0 region of the HDV but recently, sequencing and phylogenetic analysis of HDAg gene is considered a better method for genotyping. Phylogenetic analysis of nucleotide sequences of the Hepatitis Delta Antigen (HDAg) R0 region of twelve isolates from a gastroenterology treatment center in Karachi revealed presence a single genotype (HDV-1) with clustering of strains into three subgroups.

Key-words: Hepatitis D virus, hepatocellular carcinoma, phylogenetic, HDAg, interferon.

INTRODUCTION

Hepatitis delta virus (HDV) is a satellite virus that infects only in the presence of Hepatitis B virus (HBV) that can aggravate human viral hepatitis. Hepatitis Delta virus infection is the serious health problem worldwide and globally there are 15-20 million individuals reported with HDV infection, which is approximately 5% of HBV infected individuals (Botelho-Souza *et al.*, 2017). HDV has become endemic in many countries, especially in those population where HBV is in high prevalence (Rizzetto, 2015).

Hepatitis B and Hepatitis D is a confronting health issue in Pakistan, prevalence of coinfection of HDV/HBV is about 16 – 57% (Aftab *et al*., 2018). The number of individuals infected with HDV is not showing any decline. The practice of sharing contaminated needles among drug addicts, usage of non-sterile medical, surgical and dental equipment, repeated use of therapeutic injections and intravenous drips and reusing razors for shaving at barber shop are the main sources of HDV spread in Pakistan (Abbas *et al.*, 2014).

HDV is a satellite virus of HBV (Goodrum and Pelchat, 2018). Replication of HDV occurs in hepatocytes but only in the presence of HBV, which may cause cytotoxicity and immune-mediated liver damage (Negro, 2014). HBV-HDV coinfection or superinfection is more susceptible to result in hepatocellular carcinoma (HCC), compared to chronic carrier state of HBV alone, causing threefold increase in fatality (Mahale *et al*., 2019). Immunization against HBV is the only way to prevent HDV infection, while INF-α is the only proven antiviral therapy for HDV infection (Brichler *et al.*, 2018)

HDV is circular RNA negative-strand virus having a genome of 1,679 nucleotides in length, which makes it the smallest animal virus with 36nm in diameter (Goodrum and Pelchat, 2018; Opaleye *et al*., 2016). In infected cells, three types of viral RNA are known to exist; the whole genome RNA, antigenomic RNA, and the messenger RNA (mRNA) that is 900 bases in length having an antigenomic polarity. HDV has only one protein-coding gene known as Hepatitis Delta Antigen (HDAg) that is translated through this mRNA, using the host translation machinery. HDAg protein exists in two sizes; the S-HDAg having 195 amino acids (24 KDa) responsible for HDV RNA replication, and L-HDAg consisting of 214 amino acids (27 KDa) that is essential for virion assembly. Both proteins are coded by the same HDAg gene using an RNA editing mechanism where the host enzyme Adenosine Deaminase Acting on RNA1 (ADAR1) executes the cellular editing of antigenomic RNA closer to the end of the HDAg gene. This editing mutates the UAG (amber stop codon) to UGG (Tryptophan), which leads to elongation the S-HDAg, by

a carboxyl-terminal addition of 19 amino acids to form the L-HDAg protein (George *et al.*, 2014). S-HDAg is involved in genome replication while the L-HDAg, is required for virion morphogenesis (Modahl and Lai, 2000; Taylor, 2015).

HDV was previously classified into three genotypes only depending upon proportion of nucleotide distinctiveness of viral genome. After 2006, this classification was drawn out into eight clades (Chao *et al.*, 2015), where divergence between eight clades >35% and within genotype is <20% (Deny, 2006; Lempp *et al.*, 2016).

The genetic diversity of HDV is closely related to the geographic origin of the isolates. Apart from HDV-1 is common worldwide but isolates from different geographical origin are genetically distinct from each other. The European and North American regions are mostly affected with HDV-1 (Aftab *et al*., 2018), while HDV-2 and HDV-4 are mostly common in Japan, Taiwan and Russia. HDV-3 is predominately found in Amazonian region (Han *et al*., 2014), and HDV-5 to HDV-8 are found in the Central and West African regions (Nguyen *et al*., 2017). Pakistan has a high prevalence of HDV-1 (Alam *et al*., 2007).

Phylogenetic study of partial viral genomes is a convincing approach for tracing the spread of viral outbreaks. Virus phylogenetic have helped in tracing the origin and spread of Human Immunodeficiency Virus (HIV) (Scaduto *et al*., 2010) and Hepatitis C Virus (HCV) (Vandamme and Pybus, 2013). Phylogenetic analysis of delta virus HDAg helps in tracing reverse evolutionary trends, treatment management and vaccine development (Triantos *et al.*, 2012). Our previously conducted phylogenetic study revealed nucleotide variation which created subgroup within clade I (Perveen *et al.*, 2012).

In the current study, the entire HDAg coding sequence of HDV isolates from a set of HDV antibody positive patients (n=12) from a gastroenterology clinic were analyzed. The analyses aimed to investigate the evolutionary patterns and the genetic similarities and differences of HDV strains from Pakistan and assigning the local strains into the clades/genotypes of HDV. This research will help to isolate more strains from Pakistan and will provide aid for spread, control and disease management for this disease in Pakistan.

MATERIALS AND METHODS

Study Approval

The study was carried out under approval from the Ethical Review Committee of Department of Genetics, University of Karachi.

Specimen Collection

Blood samples were collected from Pakistan Health Research Council (PHRC)'s center in Jinnah Postgraduate Medical Center (JPMC), Karachi and a Gastroenterology consulting at the Doctor's Clinic. The blood samples were collected from patients who were sero-positive for Hepatitis B Virus surface antigen (HBsAg) and anti-delta antibody (HDAg) and showed presence of delta virus RNA through Polymerase Chain Reaction (PCR). Patients who are HBV and HDV positive with advanced stage of liver cirrhosis caused by HDV were also included in sampling. 5ml of whole blood was collected in purple top vacutainer with proper labelling including medical record number (MR#), lab ID, where available. Collected samples were kept on ice and then transported to Department of Genetics and stored at -30°C freezer. Clinical parameters of HDV isolates for this study is presented in Table 1.

Primer Designing

Total 43 whole genomes of HDV virus (Genotype I) available on GenBank, NCBI were aligned using Clustal Omega (McWilliam *et al*., 2013) multiple sequence alignment tool. Reverse transcription primer used for cDNA synthesis was designed from a high consensus region of the HDV genome. For whole genome sequencing the PCR primers were designed from regions of high consensus that amplified two overlapping fragments of 1093bp and 959bp, using Primer 3 software (Untergasser *et al.*, 2012). Universal M13 adapters sequences were included at 5′ of each primer to facilitate sequencing. Table 2 presents the complete information of primers.

RNA Extraction

Total RNA was extracted from 200 µL of whole blood using TRIzol® Reagent according to described protocol (Simms *et al.*, 1993). Extracted RNA was stored in a -30°C freezer.

Table 1. Clinical parameters of source patients of HDV isolates.

GenBank Accession number	Sample code	Age (Years)	Gender	ALT* (IU/L)	Disease status	HBsAg	Interferon-α response
MK890224	P-04	10	F	ND	Acute	ND	NA
MK890225	QD-01	19	M	145	Relapse	+	NR
MK890226	QD-02	50	M	45	Relapse	ND	NR
MK890227	QD-07	28	F	69	Relapse	+	NR
MK890228	QD-11	18	M	25	Acute	ND	NA
MK890229	QD-15	18	F	95	Relapse	+	NR
MK890230	QD-16	14	M	60	Relapse	+	NR
MK890231	QD-18	16	M	40	Relapse	+	NR
MK890232	QD-22	44	M	123	Acute	+	NA
MK890233	QD-34	36	F	160	Relapse	+	NR
MK890234	QD-39	28	F	68	Relapse	+	NR
MK890235	QD-52	36	M	ND	Acute	ND	NA

*: ALT: Upper normal value is less than 46 IU/L
Abbreviations: ALT: Alkaline aminotransferase, HBsAg: Hepatitis B surface antigen, ND: Not determined, NA: Not applicable, NR: Non-responder, F/M: Female/Male

Table 2. Primers used for amplification of the HDV complete genome sequences.

Primer's ID	Position in Reference[a] Genome (bp)	Orientation	Amplicon size in bp	Sequence 5' - 3'
HG-1RT	1638-1646	Antisense	-	**GATGAGCCG**
HDV_2F HDV-1R	717-736 890-911	Sense Antisense	196	**GCCGGCTGGGCAACATTCCG** **TTTCCTCTTCGGGTCGGCATGG**
M13HD1-F M13HD1-R	300-325 1400-1376	Sense Antisense	1132	**<u>GTAAAACGACCAGT</u>ACCTCCAGAGGACCCCTTCAGCGAA** **<u>CAGGAAACAGCTATGAC</u>GAGGGAGCTCCCCCGGCGAAGAG**
M13HD2-F M13HD2-R	1160-1180 430-417	Sense Antisense	984	**<u>GTAAAACGACCAGT</u>GCGGGCCGGCACTCTTCTTC** **<u>CAGGAAACAGCTATGAC</u>GAGGCTAGCCCCGTTGCTTCTTGCTTT**

[a] Reference genome GenBank accession number; NC_001653; The underscored sequence is M13 phage adapter sequences.

Reverse Transcription

Complementary DNA (cDNA) was synthesized using RevertAid® Reverse Transcriptase (MMLV-RT, cat #: EP0442, Thermo Scientific). 8 µL of extracted RNA and 2 µL of HG-1RT primer (10 µM) was incubated at 70°C for 10 minutes and then chilled on ice for 5 minutes. A 10 µl reaction mix, constituting 4 µL of 5X RevertAid®

buffer, 2 µL of 100mM of DTT, 0.5 µL of 40U/µL RiboLock RNase inhibitor (Cat #: EO0381, Thermo Scientific), 1 µL of RevertAid® reverse transcriptase (200U), 1 µL dNTPs mixture (10mM each, cat #: R0181, Thermo Scientific), and 1.5 µL water was added to the RNA-primer mix making a final volume of 20 µL. The reaction tube was incubated in a thermal cycler at 37°C for 15 minutes, 42°C for 30 minutes, 50°C for 25 minutes. The RT reaction was then terminated at 95°C for 5 minutes.

Polymerase Chain Reaction

All samples were reconfirmed for HDV RNA by amplifying an approximately 200bp fragment, using H2F and H1R primers in P-X2 thermal cycler (Thermo Electron Corporation, USA). The PCR reaction was set up using 2 µL of cDNA in 30 µL of reaction volume included with final concentration; 1.2 µL each of 10µM sense & antisense primers, 0.3 µL (5U/µL) of Thermo Taq (cat #: EP0402, Thermo Scientific), 3 µL of 10X PCR buffer, 2.4 µL of 25mM $MgCl_2$, 0.6 µL dNTPs mixture (10mM each, Thermo Scientific) and 19.3 µL of distilled water (Ref #: 10977-015, Invitrogen). Cycling conditions comprised of initial denaturation at 95°C for 2 minutes, followed by 25 cycles of denaturation at 94°C for 30 seconds, primers annealing at 66.5°C for 30 seconds, polymerization at 72°C for 40 seconds, by a final extension at 72°C for 5 minutes and a final hold at 4°C.

For whole genome amplification, two overlapping fragments were amplified by M13HD1 and M13HD2 primer's pair. 2 µL of cDNA was used in 30 µL of reaction mixture included with final concentration; 1.2 µL each of 10µM sense & antisense primers, 0.2 µL (2U/µL) of Phusion Taq polymerase (cat #: F-540S, Thermo Scientific), 6 µL of 5X PCR buffer, 0.6 µL of 50mM $MgCl_2$, 0.6 µL dNTPs (10mM each, Thermo Scientific) and 18.2 µL of distilled water (Ref #: 10977-015, Invitrogen). Cycling conditions started with initial denaturation at 98°C for 2 minutes 30 seconds, followed by 30 cycles of denaturation at 94°C for 30 seconds, primers annealing at 66.5°C for 30 seconds, polymerization at 72°C for 50 seconds, followed by a final extension at 72°C for 5 minutes and a final hold at 4°C. Gel electrophoresis was performed to resolved amplicons on a 1.5% agarose gels in 1X TAE buffer and visualized under UV light in Gel documentation system.

DNA Sequencing

All amplicons with the length of approximately 960bp and 1140bp were purified with GeneJet PCR purification kit (cat #: 0701, Thermo Scientific). Purified amplicons were sent for commercial sequencing service (Macrogen, South Korea). Each product was sequenced bidirectionally using the sense and antisense primers. Fragments amplified by primers with M13 adapters were sequenced using relevant universal M13 primers. Nucleotide sequences and chromatograms received from Macrogen electronically were analyzed using BioEdit sequence alignment editor version 7.2.6 (Hall, 1999). All sequencing chromatograms were manually screened for errors and mismatches, before assembling the forward and reverse sequencing of each amplicon. All sequences submitted to GenBank, NCBI with the accession numbers: MK890224 to MK890235, will be available online at GenBank on Nov 10, 2019.

Nucleotide Sequence Analysis

Forty-eight whole genome nucleotide sequences representing the eight clades of the delta virus, including the reference genome, NC_001653 were downloaded from GenBank for phylogenetic analysis on similar lines to the previous study (Le Gal *et al*., 2006). Whole genome HDV sequences of HDV clades (I-VIII) used in this study were as follows, for clade I: MH844625 from Pakistan, AY633627 from Iran, U81989 from Ethiopia, X04451.1 from Italy, M84917 from Lebanon, M58629 from Nauru, U81988 from Somalia, M92448 and AF104263 from Taiwan, D01075 and L22066 from USA, AB118849 from Nagasaki, X77627 from China, X85253 from Cagliari. In clade II: X60193 and AB118846 from Japan, AF104264 and U19598 from Taiwan, AJ309879 and AJ309880 from Yakutia, Russia. In clade III: L22063 from Northern South America, AB037947, AB037948 and AB037949 from Venezuela, in clade IV: AF309420, AB088679 and AB118845 from Japan, AF209859 and AF018077 from Taiwan, AB118847 from Tokyo. Clade V: AM183326 from Togo, AM183328 from Senegal, AX183331, AX741149 and AX741154; clade VI: AM183332 from Central Africa, AM183329 from Nigeria, LT594484 from Angola and AX741164; clade VII: AM183333, KM110805, KM110803 and LT604690 from Cameroon and clade VIII: AM183327 from Cote d'Ivoire, AM183330 from Senegal, GU17114 from Gabon and AX741169.

Nucleotide sequences of the R0 region comprising the region coding for the solitary HDV protein known as L-HDAg (complement to bp 1030 to 1600) in the reference genome NC_001653.1, were extracted from each of the sixty whole genome sequences. Similarly, the HDAg sequences of the twelve isolates of this study was extracted and used for phylogenetic analysis of the R0 region.

Multiple Sequence Alignment (MSA)

All sixty nucleotide sequences were aligned using Clustal Omega (Clustal Ω) (McWilliam *et al*., 2013), a web-based software available at the European Biotechnology Institute (EBI). MSA was tested and corrected for observed conflicts in the similar aligned sequences with the help of Alignment Explorer Window of the molecular evolutionary genetic analysis (MEGA X) (Kumar *et al.*, 2018) and BioEdit sequence alignment editor version 7.2.6 (Hall, 1999).

Phylogenetic Analysis

Model selection and tree construction

For selection of a nucleotide substitution model MEGA X (Kumar *et al*., 2018) was used, according to the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) the Nucleotide substitution model were decided and parameters which are most appropriate for our dataset were determined (Sullivan and Joyce, 2005).

For both AIC and BIC, the general time reversible model with gamma distribution for a proportion of invariable sites (+I) was applied to construct Maximum Likelihood (ML) tree via MEGA X. Rationality of phylogenetic tree analysis was confirmed through bootstrap (1000 pseudoreplicates).

Genetic Distance

Genetic distances were estimated by Kimura's 2 parameter method that calculated between local isolates and clade I to VIII expressed as mean ± standard deviation by using MEGA X (Kumar *et al*., 2018).

RESULTS

Phylogenetic analysis of the R0 region encompassing the HDAg using the Maximum Likelihood tree construction method using the GTR+I+G nucleotide substitution model for AIC shows that the twelve isolates from this study clustered within clade I which is the prevalent genotype in Pakistan. These isolates were found to be genotypically distinct from HDV clade II to VIII, showing maximal genotypic distance to clade III **(Fig. 1).**

Within clade I, our local isolates sub-clustered into three groups marked on the phylogenetic tree as ''this study Group-I, II, and III. Group-I is shown higher bootstrap confidence value 84, having distinct strains: MK890229, MK890233, and MK890224 are in separate branch of tree with another strain of Pakistan origin (MH844625). Group-II is shown highest bootstrap confidence value 99 and posterior probability of 100%. Group-II is associated with other geographical distinct strains: MK890231, MK890225, and MK890226 are linked together in same branch of tree. Group-III have MK890232, MK890227, MK890235, MK890230, MK890234 and MK890228 also showed separate branch in tree with bootstrap value 95 **(Fig. 1).**

DISCUSSION

In this study, we performed sequence comparisons and phylogenetic analysis of a publicly available L-HDAg sequences representing all eight clades (n=48) and 12 L-HDAg sequences of our samples. Phylogenetic analysis led to the assigning of our strains to Clade I (HDV-1), with further clustering of our sequences into distinct subgroup where the variation between subgroups was 9.37% to 14.57% **(Fig. 2),** whereas within subgroup variation for subgroup I, II, & III was 11, 3 and 6%, respectively.

These results agree with our previous phylogenetic analysis study on HDV isolates from Pakistan (Perveen *et al*., 2012). The clustering of sequences of this study isolates showing genetic heterogeneity within population which forms subgroups in clade I. Within this group, three clusters have been observed, one of which shows higher bootstrap and posterior probability values (shown as Pakistan Group II) (Fig. **1**) as compare to other two groups.

Different HDV genotypes are localized to an area or endemic to different geographical regions (Deny, 2006). The rates of genetic variation between the HDV clades are 20-40% whereas within the same clade the divergence is up to16% (Hughes *et al.*, 2011). In our study with twelve samples, the genetic variation in the HDAg region of the HDV genome showed a variation of 10%.

HDV-1 is the only clade that is ubiquitous being present all over the world, from Europe, Mediterranean, North America, Russia and Africa (Chao *et al.*, 1990; Ma *et al*., 2003; Shakil *et al*., 1997), in addition to some strains described in the Pacific island of Nauru (Chao *et al*., 1990). Previously, genotyping was carried out by RFLP analysis of the R0 region of the HDV (Ivaniushina *et al*., 2001), however, recently, sequence and phylogenetic analysis of HDAg gene or the full length HDV genome has been used *et al.*, 2015), now being considered as a more reliable method for taxonomic assignment.

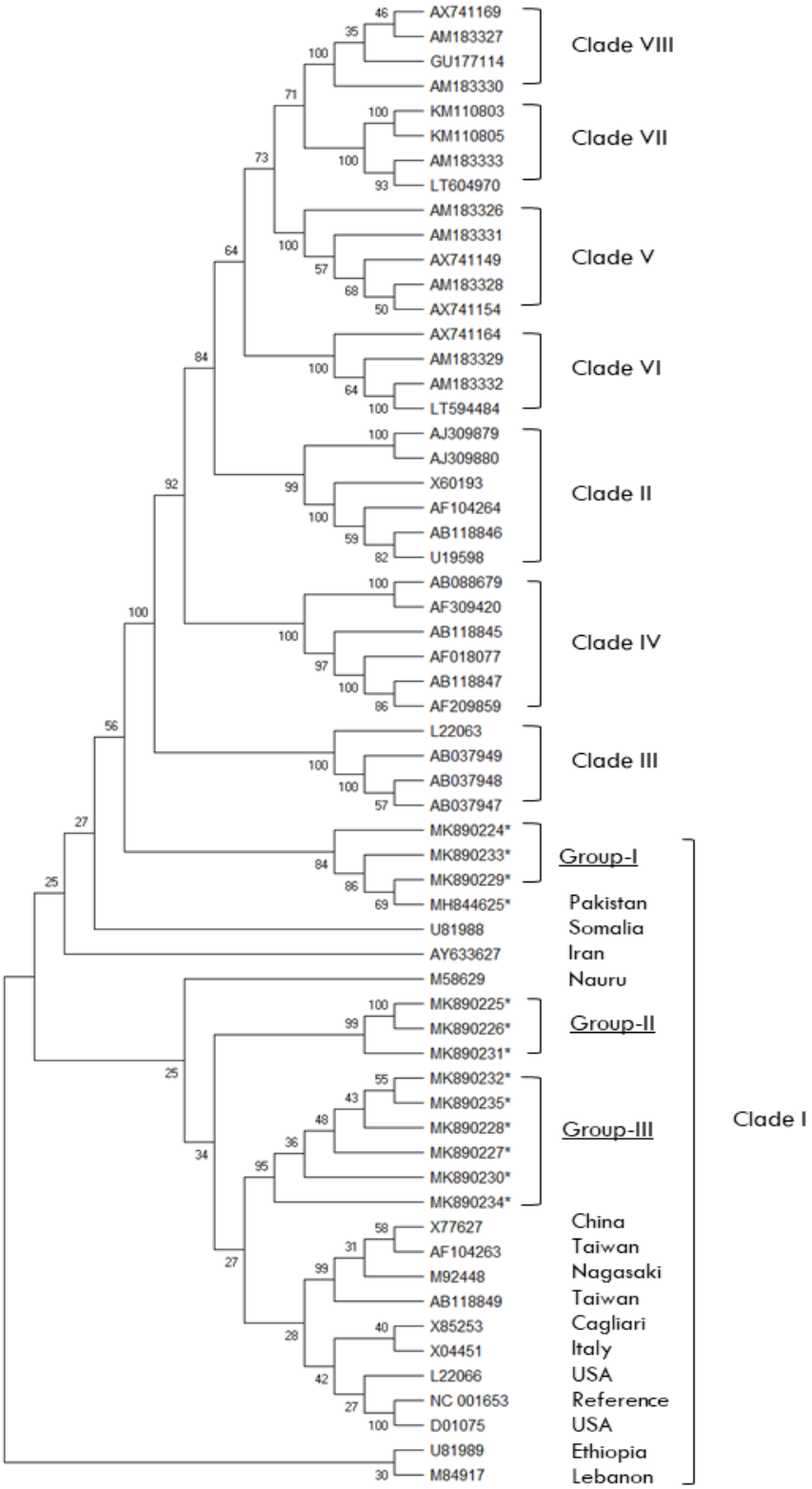

Fig. 1. Phylogenetic analysis of the HDAg coding region.

Phylogenetic analysis of the R0 region encompassing the HDAg using the Maximum Likelihood tree construction method using the General Time Reversible (GTR+I+G) nucleotide substitution model, with bootstrap consensus tree inferred from 1000

replicates. This analysis involved nucleotide sequences from all eight clades of HDV, including twelve isolates of this study. Evolutionary analyses were conducted in MEGA X.

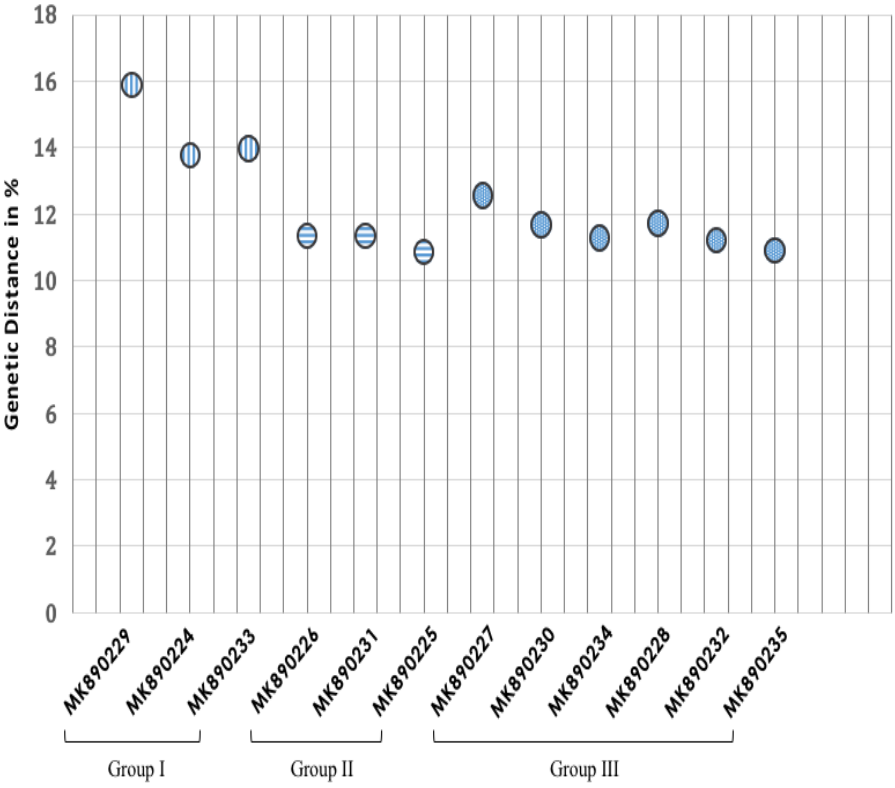

Fig. 2. Mean intra-genotypic distances of HDV isolates based on R0 region. Genetic distances in percentage (Y-axis) of twelve isolates from this study (X-axis) compared to Clade I sequences as a group. Genetic distances show formation of three subgroups.

The genetic variation among the clades is the main cause for the differences of viral pathogenesis related to different clades or genotypes. The genetic diversity in the delta virus is of concern to clinicians as it poses new complications to the available limited number of treatments for this virus (Wu *et al*., 1995). Genotyping may not help much in outlining the treatment regimen as very few effective treatments are available (Buti *et al*., 2011) however, information about the genotype of an infection helps in the prognosis of the disease (Niro *et al.*, 2011).

A high genetic diversity of the virus is also possible within an endemic area since RNA viruses have high mutation rates due to the lack of proofreading activity of the RNA polymerase during replication. The genomic changes may render the viral genome able to replicate more efficiently or to escape the host immune responses and evolve rapidly under the selection pressures present in the host. These selective pressures include the host immunity and include any drug treatments received by the patient (Wang *et al*., 2007). Therefore, an ancestral strain that landed in an area would adapt and evolve to the indigenous population's immunological and environmental pressures leading to a specific ethnogeographical clades or even subgroups within clades as we have encountered in our study.

Acknowledgement

The study was funded by the Higher Education Commission (HEC), National Research Program for Universities (NRPU), Project No.20-4473/NRPU /R&D/HEC/14/1200, granted to Dr. Obaid Y. Khan.

REFERENCES

Abbas, Z., G.N. Soomro, S. M. Hassan and N.H. Luck (2014). Clinical presentation of hepatitis D in Pakistani children. *Eur J Gastroenterol Hepatol.*, 26(10): 1098-1103.

Aftab, M., I. Amin, M. Idrees, A. Ali, S. Rafique and S. Naz (2018). Molecular epidemiology of hepatitis delta and hepatitis B viruses circulating in two major provinces (East and North-West) of Pakistan. *Infect Genet Evol.*, 64: 65-69.

Alam, M. M., S.Z. Zaidi, S. A. Malik, S. Shaukat, A. Naeem, S. Sharif and J.A. Butt (2007). Molecular epidemiology of Hepatitis B virus genotypes in Pakistan. *BMC Infect Dis.,* 7(1): 115.

Botelho-Souza, L. F., D. Souza Vieira, A. de Oliveira Dos Santos, A.V. Cunha Pereira and J.M. Villalobos-Salcedo (2015). Characterization of the Genotypic Profile of Hepatitis Delta Virus: Isolation of HDV Genotype-1 in the Western Amazon Region of Brazil. *Intervirology,* 58(3): 166-171.

Botelho-Souza, L. F., M. P. A., Vasconcelos, A.O. Dos Santos, J. M. V. Salcedo and D.S. Vieira (2017). Hepatitis delta: virological and clinical aspects. *Virol J.,* 14(1): 177.

Brichler, S., M. Setshedi and C. Renou (2013). Resolution of chronic hepatitis delta infection after five years of peginterferon-adefovir: lessons from a case report. *Clin Res Hepatol Gastroenterol,* 37(3): e81-84.

Buti, M., M. Homs, F. Rodriguez-Frias, G. Funalleras, R. Jardi, S. Sauleda and R. Esteban (2011). Clinical outcome of acute and chronic hepatitis delta over time: a long-term follow-up study. *J Viral Hepat.,* 18(6): 434-442.

Chao, M., C.C. Lin, F.M. Lin, H.P. Li and S.B. Iang (2015). Whole-genome analysis of genetic recombination of hepatitis delta virus: molecular domain in delta antigen determining trans-activating efficiency. *Journal of General Virology,* 96(12): 3460-3469.

Chao, Y. C., M.F. Chang, I. Gust and M.M. Lai (1990). Sequence conservation and divergence of hepatitis delta virus RNA. *Virology,* 178(2): 384-392.

Deny, P. (2006). Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? *Curr Top Microbiol Immunol.,* 307: 151-171.

Farci, P. and G. Anna Niro, G. (2018). Current and Future Management of Chronic Hepatitis D. *Gastroenterology & hepatology,* 14(6): 342-351.

George, C. X., L. John and C.E. Samuel (2014). An RNA editor, adenosine deaminase acting on double-stranded RNA (ADAR1). *J Interferon Cytokine Res.,* 34(6): 437-446.

Goodrum, G. and M. Pelchat (2018). Insight into the Contribution and Disruption of Host Processes during HDV Replication. *Viruses,* 11(1): 21.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.

Han, M., M. Littlejohn, L. Yuen, R. Edwards, U. Devi, S. Bowden and K. Jackson (2014). Molecular epidemiology of hepatitis delta virus in the Western Pacific region. *J Clin Virol., 61*(1), 34-39.

Hughes, S. A., H. Wedemeyer and P.M. Harrison (2011). Hepatitis delta virus. *Lancet,* 378(9785): 73-85.

Ivaniushina, V., N. Radjef, M. Alexeeva, E. Gault, S. Semenov, M. Salhi and P. Deny (2001). Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia. *J Gen Virol.,* 82(Pt 11): 2709-2718.

Kumar, S., G. Stecher, M. Li, C Knyaz and K, Tamura (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution,* 35(6): 1547-1549.

Le Gal, F., E. Gault, M. Ripault, J. Serpaggi, J. Trinchet, E. Gordien and P. Dény (2006). Eighth Major Clade for Hepatitis Delta Virus. *Emerging Infectious Diseases,* 12(9): 1447-1450.

Lempp, F. A., Y. Ni and S. Urban (2016). Hepatitis delta virus: insights into a peculiar pathogen and novel treatment options. *Nat Rev Gastroenterol Hepatol.,* 13(10): 580-589.

Ma, S. P., H. Sakugawa, Y. Makino, M. Tadano, F. Kinjo and A. Saito (2003). The complete genomic sequence of hepatitis delta virus genotype IIb prevalent in Okinawa, Japan. *J Gen Virol.,* 84(Pt 2): 461-464.

Mahale, P., P. Aka, X. Chen, R.M. Pfeiffer, P. Liu, S. Groover and T.R. O'Brien (2019). Hepatitis D virus infection, cirrhosis and hepatocellular carcinoma in The Gambia. *J Viral Hepat*., doi:10.1111/jvh.13065

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y.M. Park, N., Buso and R. Lopez (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research,* 41(W1): W597-W600.

Modahl, L. E. and M.M. Lai (2000). The large delta antigen of hepatitis delta virus potently inhibits genomic but not antigenomic RNA synthesis: a mechanism enabling initiation of viral replication. *J Virol.,* 74(16): 7375-7380.

Negro, F. (2014). Hepatitis D virus coinfection and superinfection. *Cold Spring Harb Perspect Med.,* 4(11): a021550. doi:10.1101/cshperspect.a021550

Nguyen, H. M., B.T. Sy, N. T. Trung, N. X. Hoan, H. Wedemeyer, T.P. Velavan and C.T. Bock (2017). Prevalence and genotype distribution of hepatitis delta virus among chronic hepatitis B carriers in Central Vietnam. *PLoS One,* 12(4): e0175304. doi:10.1371/journal.pone.0175304

Niro, G. A., D. Gioffreda and R. Fontana (2011). Hepatitis delta virus infection: Open issues. *Digestive and Liver Disease,* 43: S19-S24.

Opaleye, O. O., O.M. Japhet, O. M. Adewumi, E. C. Omoruyi, O. A. Akanbi, A. S. Oluremi and C.T. Bock (2016). Molecular epidemiology of hepatitis D virus circulating in Southwestern Nigeria. *Virology Journal,* 13: 61-61.

Perveen, S., M.I. Nasir, S.M. Shahid, A. Azhar and O.Y. Khan (2012). Phylogenetic analysis of HDV isolates from HBsAg positive patients in Karachi, Pakistan. *Virol J.,* 9: 162.

Rizzetto, M. (2015). Hepatitis D Virus: Introduction and Epidemiology. *Cold Spring Harbor perspectives in medicine,* 5(7): a021576-a021576.

Scaduto, D. I., J.M. Brown, W. C. Haaland, D. J. Zwickl, D. M. Hillis and M.L. Metzker (2010). Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences. *Proceedings of the National Academy of Sciences,* 107(50): 21242-21247.

Shakil, A. O., S. Hadziyannis, J.H. Hoofnagle, A. M. Di Bisceglie, J. L. Gerin and J.L. Casey (1997). Geographic distribution and genetic variability of hepatitis delta virus genotype I. *Virology,* 234(1): 160-167.

Simms, D., P.E. Cizdziel and P. Chomczynski (1993). TRIzol: A new reagent for optimal single-step isolation of RNA. *Focus,* 15(4): 532-535.

Sullivan, J. and P. Joyce (2005). Model Selection in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics,* 36(1): 445-466.

Taylor, J. M. (2015). Hepatitis D Virus Replication. *Cold Spring Harbor perspectives in medicine,* 5(11): a021568.

Triantos, C., M. Kalafateli, V. Nikolopoulou and A. Burroughs (2012). Meta-analysis: antiviral treatment for hepatitis D. *Aliment Pharmacol Ther.,* 35(6): 663-673.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm and S.G. Rozen (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.,* 40(15): e115.

Vandamme, A.-M. and O.G. Pybus (2013). Viral phylogeny in court: the unusual case of the Valencian anesthetist. *BMC Biology,* 11(1): 83. doi:10.1186/1741-7007-11-83

Wang, S. Y., J.C. Wu, T. Y. Chiang, Y.H. Huang, C. W. Su and I.J. Sheen (2007). Positive selection of hepatitis delta antigen in chronic hepatitis D patients. *J Virol.,* 81(9): 4438-4444.

Wu, J. C., K.B. Choo, C. M. Chen, T. Z. Chen, T. I. Huo and S.D. Lee (1995). Genotyping of hepatitis D virus by restriction-fragment length polymorphism and relation to outcome of hepatitis D. *Lancet,* 346(8980): 939-941.

(*Accepted for publication June 2019*)