Molecular characterization of hepatitis A virus strains in a tertiary care health set up in north western India

Mini Pritam Singh, Manasi Majumdar, Babu Ram Thapa*, Puneet Kumar Gupta, Jasmine Khurana, Bimal Budhathoki & Radha Kanta Ratho

Departments of Virology & *Pediatric Gastroenterology, Postgraduate Institute of Medical Education & Research, Chandigarh, India

Received May 27, 2013

**Background & objectives:** Hepatitis A virus usually causes acute viral hepatitis (AVH) in the paediatric age group with a recent shift in age distribution and disease manifestations like acute liver failure (ALF). This has been attributed to mutations in 5’non-translated region (5’NTR) which affects the viral multiplication. The present study was aimed to carry out the molecular detection and phylogenetic analysis of hepatitis A virus strains circulating in north western India.

**Methods:** Serum samples from in patients and those attending out patient department of Pediatric Gastroenterology in a tertiary care hospital in north India during 2007-2011 with clinically suspected AVH were tested for anti-hepatitis A virus (HAV) IgM antibodies. Acute phase serum samples were subjected to nested PCR targeting the 5’NTR region followed by sequencing of the representative strains.

**Results:** A total of 1334 samples were tested, 290 (21.7%) were positive for anti-HAV IgM antibody. Of these, 78 serum samples (< 7 days old) were subjected to PCR and 47.4% (37/78) samples showed the presence of HAV RNA. Children < 15 yr of age accounted for majority (94%) of cases with highest seropositivity during rainy season. Sequencing of 15 representative strains was carried out and the circulating genotype was found to be III A. The nucleotide sequences showed high homology among the strains with a variation ranging from 0.1-1 per cent over the years. An important substitution of g to A at 324 position was shown by both AVH and ALF strains. The cumulative substitution in AVH strains Vs ALF strains as compared to GBM, Indian and prototype strain in the 200-500 region of 5’NTR was comparable.

**Interpretation & conclusion:** Our results showed hepatitis A still a disease of children with III A as a circulating genotype in this region. The mutations at 5’NTR region warrant further analysis as these affect the structure of internal ribosomal entry site which is important for viral replication.

**Key words** Genotype - hepatitis A virus - mutations - phylogenetic analysis - serology
Hepatitis A virus (HAV) accounts for 60–80 per cent cases of clinical hepatitis in the developing countries. The clinical course may vary from asymptomatic, acute self-limiting viral hepatitis to life threatening acute liver failure (ALF). In 90 per cent of the children, the infection is subclinical followed by seroconversion leading to a life-long immunity. It has been seen that severe manifestations are more common in young adults requiring hospitalization in 22.3 to 52.4 per cent cases with overall case fatality rate of 0.3 per cent. The changing environmental, socio-economic conditions and the availability of a commercially available vaccine have markedly affected the epidemiology of the disease causing a gradual shift in the age distribution to 2nd and 3rd decade of life.

HAV is the only member of the genus Hepatovirus in the family Picornaviridae causing human infection. HAV is known to display a high degree of antigenic and genetic conservation contrary to the high frequency of genetic changes seen in RNA viruses. Molecular epidemiology of HAV is important to understand the strains circulating in various geographical regions and tracing the source of contamination in an outbreak situation. The HAV strains isolated from various parts of the world constitute a single serotype and are divided into six genotypes (I–VI). Genotypes I-III are most commonly associated with human infections and have a variable geographical distribution. Majority of human strains (80%) belong to genotype I. The predominantly circulating genotype in India is genotype III. However, a few studies have reported circulation of genotype IA in New Delhi and also co-circulation of genotypes IIIA and IB has been reported from a day care center in Pune, western India. The molecular characterization of the infectious agents is important, as it provides the information about the circulating strains in a particular region, the invasion of new strains from different geographical areas and their role in the pathogenesis and severity of the disease.

The aim of this study was to carry out the molecular characterization of the prevalent strains of HAV over a period of four years. This study was carried out in a tertiary care hospital of north west India which caters Chandigarh and the adjoining States of Haryana, Punjab, Himachal Pradesh, Jammu and Kashmir, parts of Uttar Pradesh and Rajasthan.

**Material & Methods**

The blood samples were received in the department of Virology from patients with clinically suspected viral hepatitis from March 2007 to August 2011 visiting the in- and out-patients of Pediatric Gastroenterology department of the Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh. The samples in 2009 could not be tested due to the non-availability of ELISA kits during this time. The blood samples were collected and transported in cold chain system for the detection of anti HAV IgM antibodies. The study protocol was approved by the institute’s ethical committee. A total of 1334 clotted blood samples were received, the serum was separated, and the vials were coded and stored at -70 °C in aliquots till tested.

**Case definition:** The clinical details were available for some of the patients who were admitted with either acute viral hepatitis (AVH) or acute liver failure (ALF). The AVH was defined as the patients presenting with serum aspartate aminotransferase (AST) or alanine aminotransferase (ALT) elevation of at least five-fold with clinical jaundice and without evidence of any chronic liver disease. ALF was defined as biochemical evidence of liver injury, no history of known chronic liver diseases, coagulopathy not corrected by vitamin K administration, international normalized ratio (INR) >1.5 if the patient had encephalopathy or >2.0 if the patient did not have encephalopathy.

**Serology:** The serum samples were tested for anti-HAV IgM antibodies (Immunovision, USA) using commercially available IgM capture ELISA kit with a sensitivity and specificity of >99 per cent as per the manufacturer’s instructions.

**RNA extraction and reverse transcription-polymerase chain reaction (RT-PCR) for detection of HAV genome:** Acute phase serum samples (< 7 days old, n=78) positive for anti-HAV IgM antibodies were subjected to nested PCR targeting the 5’NTR (5’-non translated region). Briefly, 140 µl of serum was used for RNA extraction using QIAamp Viral RNA Mini Kit (Qiagen, Germany). The RNA was eluted in 50 µl of the elution buffer and used as a template for cDNA synthesis. On the same day, 10 µl of the dissolved RNA was reverse transcribed to cDNA using external anti-sense primers (Sigma, USA) and RevertAid™ First Strand cDNA Synthesis Kit (MBI Fermentas, USA). Five µl of the cDNA produced was used as a template for the first round of nested PCR. Briefly, the final concentration of various reagents used in a reaction volume of 25 µl was; 1X reaction buffer, 0.2 mM dNTP mix, 1.5 mM MgCl2, 0.5 µM of each outer sense and outer antisense.
and 0.6 U/µl of Taq polymerase. The thermal profile of PCR amplification cycle was: initial denaturation at 94°C for 3 min followed by 35 cycles of denaturation at 95°C for one min; annealing at 50°C for one min and extension at 72°C for seven min. Five µl of amplicons of first round of PCR were used as template for second round of nested PCR. The inner sense and antisense primers were used in second round of nested PCR keeping the thermal profile unchanged. Positive and negative controls were included with each run and all necessary precautions were taken to prevent cross-contamination. Ten µl of second round amplicons were electrophoresed in two per cent agarose gel containing ethidium bromide and examined for specific band of 563 bp under ultraviolet light. Representative second round amplicons of nested PCR were column purified (QIA gel purification kit, Qiagen, Germany) and both the forward and reverse strands were sequenced for 15 representative strains (10 AVH and all the 5 ALF patients) using Big Dye Terminator cycle sequencing Ready Reaction Kit (Applied Biosystems, USA) and an automated sequencer (ABI PRISM 3100 Genetic Analyzer, Applied Biosystems, USA).

**Sequence analysis and phylogenetic tree construction:**

The phylogenetic status of 15 representative 5’NTR positive specimens covering all years of study was analyzed. ABI chromatogram files were viewed using Finch TV 1.4.0, (Geospiza, Inc., USA) following sequence drafting with bioedit 7.0.9 software. For the confirmation of these sequences, database search was implemented using BLAST program available at NCBI website (http:// blast.ncbi.nlm.nih.gov/blast.cgi). For sequence comparison, standard representative strains sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank). Clustal X 2.0.11 programme (http://www.clustal.org/clustal2/) was used for multiple sequence alignment followed by Molecular Evolutionary Genetics Analysis (MEGA4) (Tempe, USA) for phylogenetic tree construction.

Accession numbers and designations of the complete or partial genome sequence employed for the present study were as follows:

- Genotype IA: A0020564, AB020567, AB020566, AB020569, AF485328, AF512536, X75215(GBM/WT)
- Genotype IB: M 20273, AF268396, M 14707, AF314208
- Genotype IIA: Ay644670
- Genotype IIB: Ay644676
- Genotype IIIA: AJ299464, AY644337, AB279732, AB279733, AB279734, DQ991029, FJ360734.1
- Genotype IIIB: AB258387, AB279735
- Genotype VI: D00924.1

**Statistical analysis:**

The statistical analysis was carried out using Statistical Package to Social Sciences (SPSS Inc.Chicago, IL, version 15.0 for windows) and P<0.05 was considered significant.

**Results**

The mean age group of the study subjects was 8.23 ± 9.45 yr with 95 % confidence interval ranging from 5.98 to 10.47 yr and the male: female ratio was 2.6:1. Majority of patients (94%) were below 15 yr of age. Nearly 70 per cent of patients were hospitalized. The seasonal distribution of anti-HAV IgM positivity over the years is shown in Fig. 1.

Of the 1334 samples tested, 290 (21.7%) were positive for anti-HAV IgM antibody. Of these, 78 acute serum samples (< 7 days old) were subjected to PCR targeting the 5’NTR of HAV RNA, and 37 (47.4%) samples showed presence of HAV RNA.

**Phylogenetic analysis of HAV isolates:**

A total of 15 nucleotide sequences from this study have been deposited in the GenBank under the accession numbers from JX390729-JX390733 and JX481901-JX481910 (Table). Phylogenetic analysis was performed using neighbor joining method and boot strap values of 1000 replicates were used for statistical verification; >85 per cent similarity has been reported between various genotypes of hepatitis A. The clustering topology of the phylogram revealed that the HAV strains circulating in our geographical region (2007-2011) belonged to genotype III, subtype IIIA. The sequences obtained over the years showed high level of sequence homology with sequence variability ranging between 0.1-1 per cent. Sequences from AVH and ALF cases showed 97-98 per cent nucleotide sequence identity when compared with the genotype- IIIA prototype strain originally isolated from Panamanian owl monkeys (PA21, accession no. M63026.1).

Seven of our strains (JX481901, JX481902, JX481904, JX481906, JX481908, JX390730 and JX390731) clustered along the Indian isolate of HAV (FJ 360734.1) reported from Pune, western India. One of our ALF sequence (JX390732) showed close similarity to the wild type strain of HAV genotype IIIA reported from blood donation camp in Germany AY644337. Three of our sequences (JX481905, JX481910, and JX390729) showed close association with a reported HAV strain from a Guillain-Barre patient with longer duration of viraemia (DQ991029). Another three sequences (JX390733, JX481907, and JX481909)
Fig. 1. Seasonal distribution of anti-HAV IgM positivity during 2007-2011 (Anti-HAV IgM seropositivity was highest in the rainy season i.e 41.4, 53.8 and 56.2 per cent in the year 2007, 2010 and 2011, respectively, except in the year 2008 where the maximum seropositivity of anti-HAV IgM was observed in the post-monsoon period). The mean ± standard deviation (SD) for anti-HAV IgM seropositivity and lower-upper 95 per cent confidence interval (CI) over the years are represented below the bar graph.

Table. GenBank accession numbers, genotype and year of isolation of HAV strains reported from north India

<table>
<thead>
<tr>
<th>GenBank accession no.</th>
<th>Strain code</th>
<th>Sub-genotype</th>
<th>Year of isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>JX481901</td>
<td>HAV 177</td>
<td>III A</td>
<td>2007</td>
</tr>
<tr>
<td>JX481902</td>
<td>HAV 3332</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX481903</td>
<td>HAV 4000</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX481904</td>
<td>HAV A1</td>
<td>III A</td>
<td>2007</td>
</tr>
<tr>
<td>JX481905</td>
<td>HAV A2</td>
<td>III A</td>
<td>2009</td>
</tr>
<tr>
<td>JX481906</td>
<td>HAV A3</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX481907</td>
<td>HAV A5</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX481908</td>
<td>HAV A6</td>
<td>III A</td>
<td>2009</td>
</tr>
<tr>
<td>JX481909</td>
<td>HAV A9</td>
<td>III A</td>
<td>2010</td>
</tr>
<tr>
<td>JX481910</td>
<td>HAV A10</td>
<td>III A</td>
<td>2010</td>
</tr>
<tr>
<td>JX390729</td>
<td>HAV 347</td>
<td>III A</td>
<td>2007</td>
</tr>
<tr>
<td>JX390730</td>
<td>HAV 3328</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX390731</td>
<td>HAV 3361</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX390732</td>
<td>HAV A4</td>
<td>III A</td>
<td>2011</td>
</tr>
<tr>
<td>JX390733</td>
<td>HAV 2415</td>
<td>III A</td>
<td>2008</td>
</tr>
</tbody>
</table>

were related to HAV isolate (AJ299464) reported from a HAV outbreak in Norway associated with parental transmission among hemophiliacs and intravenous drug users. One of our strains (JX481903) was related to a strain isolated from Japan AB279732. (Fig. 2).

Substitutions between nucleotides 200 and 500 of the 5’NTR region: Interestingly all our strains showed an important substitution of G to A at nucleotide 324 position when compared with wild-type HAV GBM/WT RNA (X75215). As proposed by Fujiwara et al, this nucleotide is positioned at the beginning of the stem structure of domain IV, and the change increases the free energy resulting in more stable structure. However, no difference was found in distribution of this substitution related to disease severity. On comparison of strains isolated from our region with wild-type HAV GBM/WT RNA (X75215), prototype strain for genotype IIIA (M63026.1) and complete genome sequence of an Indian strain reported from Pune, Maharashtra (FJ360734.1 IND-HAV-99F) nucleotide identities were found to range from 90 to 92 per cent, 97 to 98 per cent
and 98 to 99 per cent, respectively. HAV GBM strain was recovered from a human faecal sample collected in the pre-icteric phase of infection in Germany and was named GBM wild type as was adapted for rapid growth and release in HFS (human lung fibroblast cells) and, FRhk-4 cells (Foetal Rhesus monkey kidney-derived cells)\textsuperscript{21}. Genotype circulating in north west India was found to be II IA. Therefore, the strains were next compared with PA21 which is the prototype strain of HAV genotype II IA and was originally isolated from Panamanian owl monkey, but has subsequently been found associated with human cases of HAV from Nepal and Northern India during 1989 and 1990\textsuperscript{12}. Further, we tried comparing the strain reported in the present study from north west with an indigenous strain. Whole genome sequence of HAV genotype III A was available only from western India and, therefore, IND-HAV-99F was selected, which was isolated from stool sample of an icteric patient from western India in 1999\textsuperscript{22}. The overall nucleotide substitution in 200-500 nt region as compared to the above mentioned standard strains are represented in Fig. 3a, b and c. Further comparison of our strains with IND-HAV-99F revealed a 161 A to G substitution in 5 of 15 (33.3\%) cases, 207 C to T in 8 of 15 (53.3\%) cases and 559 A to G in 3 of 15 (20\%) cases. On comparison with prototype strain of genotype III A (PA21) five substitutions were observed in domain III of IRES 105 C to T in 15 of 15 (100\%) cases, 148 T to C in 6 of 15 (40\%), 187 C to T in 15 of 15 (100\%), 213 G to A in 15 of 15 (100\%), 266 G to A in 15 of 15 (100\%).
In domain IV two substitution and one insertion were observed 490G to A in 11 of 15 (73.3%), 517T to C in 15 of 15 (100%) and at position of 431 insertion of G was observed in all the strains.

On comparison with GBM/WT (X75215) AVH cases had significantly higher number of substitution than ALF cases (14.50 ± 0.97 vs 13.40 ± 0.55; P=0.0369). On comparing with FJ360734.1 Indian strain a higher mean substitution in AVH patients was observed when compared to ALF but the difference was not statistically significant (1.4 ± 1.17 vs 1.0 ± 0.71; P=0.5000). Similar was the observation on comparing AVH and ALF strains with HAV genotype IIIA prototype strain PA21 (M63026.1) (4.4 ± 1.17 vs 3.6 ± 0.55; P=0.1769).

Discussion

Hepatitis A occurs worldwide, however geographic differences exist in its endemicity23. In highly endemic areas like in Asia, Africa, Latin America, and the Middle East, the prevalence of HAV IgG antibodies reaches 90 per cent in adults, and most children have been infected by 10 yr of age23,24. In our study, the predominant age groups affected continued to be children less than 15 yr of age which accounted for 94 per cent of all cases which is in concordance with previous reports from India23,24-26. However, a few studies from India have shown a shift in the epidemiology of hepatitis A to older age group, based on the serological surveys1,27. This was because the children belonging to lower socio-economic status excrete HAV in large quantities and a substantial pool of susceptible (anti-HAV negative) adolescents/adults is present among the higher socio-economic class. The molecular detection of HAV RNA ranged from 39.7 per cent in a hospital based study15 to 63.2 per cent in outbreak investigations8. We got a comparable result of 47.4 per cent (37/78).

Using molecular approach two distinct transmission patterns of HAV have been established: (i) infection with indigenous strain, and (ii) infection due to imported strain. Robertson et al7 reported that HAV strains from USA formed a discrete, closely related geographical cluster within genotype IA suggesting existence of circulating endemic population of HAV. In contrast strains collected from Western Europe (Sweden, Germany, Denmark and Switzerland) differed widely in their respective genotypes suggesting little endemic transmission and most cases were related to importation from other endemic regions7. Since, HAV is mostly a self-limiting disease of tropics; data regarding currently circulating strains are limited. In this study, the representative strains collected over the years (2007-2011) from north India were sequenced and found to be genotype IIIA. HAV genotype IIIA has been reported as an aetiological agent of various other water borne outbreaks from northern5, southern10 and western11 India. Hussain et al28 have reported genotype III as the predominant genotype (70%) followed by genotype IA (30%) in Delhi. Co-circulation and co-infections with subgenotypes IIIA and IB have been reported from Pune, western India13. However, none of our isolates belonged to genotype IA and hence we presumed that it was genotype IIIA which was predominantly circulating in our region. In contrast, majority of the circulating strains reported from the adjoining country i.e. China over the year 2003-2008 belonged to genotype IA29.

On evaluating the genetic diversity in the 5’NTR region, it was found that all our strains had 99-99.9 per cent nucleotide identity (PNI) with each other,
The variation in the severity of hepatitis A may be attributed to various viral and host factors. Among the viral factors, the proposed factors are duration of viraemia, viral load and unique amino acid substitutions in hepatitis A virus strains, leading to alterations in the surface charge distribution on the VP1 protein. This contributes to neutralization escape of virus leading to a longer duration of viraemia. Further, association between severity of hepatitis A and nucleotide substitution in 5′NTR region has also been postulated. Fujiwara et al. reported whole genome analysis of hepatitis A virus in three cases with AVH and showed possible association between the severity of hepatitis A and nucleotide substitution in 5′NTR region. They also reported that the mutations in 5′NTR were associated with cytopathic variants in cultured cell, and virulence in tamarin. The 5′NTR region forms a highly ordered secondary structure and contains elements necessary for RNA translation and viral replication. Schultz and colleagues have reported that mutations within 5′NTR of cell adapted HAV enhance cap-independent translation directed by HAV internal ribosomal entry site (IRES) in a cell type specific fashion.

Both the AVH and ALF strains reported from our region showed an important substitution of G to A at nucleotide 324 at the beginning of stem structure of domain IV, this change increases the free energy resulting in more stable structure as proposed by Fujiwara et al. These strains share high nucleotide homology with each other as well as with strains reported from western (FJ360734.1) and southern India (DQ004690). Hence, the 5′NTR region of hepatitis A virus is conserved and may represent a viable antiviral agent. On comparing the mutations in the IRES region with respect to disease severity the strains from AVH patients had higher substitution compared to strains from ALF patients, though the difference was not significant. However, a limitation of our study was that sequences from only five ALF patients were analyzed.

The present findings showed the predominance of HAV genotype IIIA circulating in north west India from 2007-2011. However, other genotypes with limited circulation might be missed because of the restricted number of samples sequenced. Therefore, continuous surveillance is required to get a comprehensive overview of the circulating genotypes. On evaluating the per cent nucleotide identity it was observed that the 5′NTR region of the strains circulating in north west India was highly conserved. Evaluation of the nucleotide substitutions of the strains depending on disease severity i.e AVH versus ALF showed higher number of substitutions in AVH strains than ALF strains but the difference was non significant. Therefore, further pathogenicity studies involving larger number of samples need to be carried out to understand the functional aspect of these mutations. In the present study we were not able to perform the viral load, hence future studies are required to see correlation of viral load with mutation and the disease severity.

Acknowledgment

The authors thank Shriyut Om Prabhash, Junior laboratory technician and B.M. Dhawan, Senior laboratory technician of the Department of Virology for performing the serological tests.

References


Reprint requests: Dr Mini Pritam Singh, Department of Virology, Postgraduate Institute of Medical Education & Research, Chandigarh 160 012, India

E-mail: minipsingh@gmail.com