Abstract

Hepatitis A is a major public health problem that causes significant morbidity and mortality all over the world. Approximately 1.5 million cases of disease with overt symptoms are reported annually worldwide. In India, transition from hyperendemic to intermediate endemic status of hepatitis A can be expected to result in disease patterns characterized by increased morbidity, increased disease burden and large community outbreaks thereby increasing healthcare cost. Molecular characterization studies on HAV mainly include genotyping methods that analyze short genome fragments. Full-length genome sequencing has been shown to be essential due to limitations of genotyping methods in identifying the antigenic variants and classifying the strains clearly. This strategy is also recommended for determination of intra and intertypic recombination of HAV. Globally, molecular characterization of HAV represents mainly the strains from low endemic regions rendering the strains from endemic regions underreported. Limited studies carried out to characterize HAV strains from India have shown cocirculation of genotype IA, IB along with predominance of genotype IIIA. Genotype IIIA has re-emerged in Europe and is becoming more prevalent than formerly assumed. Greater genetic variability within genotype IIIA than within genotype IA strains which are predominant globally has been suggested. Characterization of full-length genomes becomes indispensable in India due to cocirculation of more than one genotype.

Present study describes characterization of full-length genomes of HAV genotype IIIA strains from Indian subcontinent for the first time and deals with three HAV strains, two from acute infection (CP-IND and PN-IND) and one associated with Guillain-Barre Syndrome (GBS) subtype acute motor axonal neuropathy (AMAN).

Full-length nucleotide and amino acid sequence data of Indian strains were compared with that of the GenBank genotype IIIA strains of HAV. Percent nucleotide identities (PNI) were in the range of 94.76 to 97.25 for CP-IND strain, 94.87 to 97.83 for PN-IND and 94.35 to 97.83 for GBS-IND strain. Percent amino acid differences (PAAD) were in the range of 0.14 to 0.4 for CP-IND, 0.14 to 0.49 for PN-IND strain and 0.4 to 0.72 for that of GBS-IND strain. Within three Indian strains PNI was in the range of 96.60 to 97.83. PAAD was 0.18 when Indian strains CP-IND and PN-IND were compared with each other. GBS-IND strain showed PAAD of 0.4 and 0.49 with CP-IND and PN-IND strains respectively.

Comparison of Indian sequences with sequences of other genotypes at full-length level and at P1, P2 and P3 domain level showed PNI in the range of 80-83. However, Percent Amino Acid Identity (PAAI) values were in the range of 96-97 at P1
domain, 92-93 and 91-92 at P2 and P3 domains respectively among human strains. The entire ORF among isolates of different genotypes and subgenotypes was well conserved, at amino acid level differing by only up to 8.14% and 2% respectively. 5' NCR of all HAV strains of genotype IIIA from India, Norway, Germany and Japan appears to be 715-716nt in length due to 18 common deletions and 1 addition as compared to that of HM175/wild type strain.

Analysis of synonymous (Ks) and nonsynonymous (Ka) mutations indicated very high Ks/Ka ratios for GenBank IIIA sequences than that of two Indian strains (CP-IND and PN-IND) indicating higher number of nonsynonymous changes in the later. GBS-IND strain was analyzed separately due to its association with different clinical condition of disease. This strain showed higher number of nonsynonymous mutations as compared to other IIIA strains.

Phylogenetic analyses of all positions and only nonsynonymous nucleotide positions of HAV strains available till date representing most of the genotypes (except VI) showed a branch for simian HAV strains (IV and V) separate from genotypes found in human hepatitis A infections (I, II, and III) with bootstrap support of 70% and 97% respectively indicating host specific evolution of HAV strains.

HAV strain associated with GBS showed higher number of nucleotide and amino acid substitutions as compared to each of the other genotype IIIA strains including Indian strains. Among amino acid substitutions, seven were unique to GBS-IND strain as compared to all HAV strains deposited in GenBank. Majority of these substitutions were heterologous and involved in or close to predicted B cell epitope regions. Homology modeling approach applied to capsid proteins of HAV strains indicated a localized change in VP1 capsid protein of GBS-IND strain from positive and basic to negative and acidic suggesting a possibility of change in receptor binding specificity. Unique and heterologous amino acid substitutions (n=3) of GBS-IND strain were also detected within highly conserved amino acid stretches suggesting possible potential of viral genomic mutations to alter protein functions and clinical features of the disease.

Evaluation of different genomic regions of HAV using various phylogenetic methods confirmed the utility of 5'NCR and VP4 region in diagnostics due to their conserved nature and identified VP3, 2A, 2C and 3D as suitable genomic regions comparable to VP1 region used as a gold standard for genotyping. Likelihood Mapping analysis indicated complete genome sequence as the most suitable choice for HAV genotyping that was followed by 2C region, which has not been explored till date.