Prevalence of hepatitis B virus genotypes and sub-genotypes in north and east regions of India: DNA sequencing methodology

Jagdish C. Kandpal1*, Rishendra Kumar1, Kailash Chandra2, Santosh Pandey 3, Tara S. Bisht1
1 Department of Biotechnology, Kumaun University Nainital, Uttarakhind, India, 2 Department of Biochemistry, HIMSR, JamiaHamdard, Delhi, India, 3 Department of Molecular Biology, CORE Diagnostics, Gurugram, Haryana, India

Corresponding author : jagdish.kandpal@yahoo.com

Abstract

Introduction: Hepatitis B virus (HBV) is a highly prevalent infecting virus among liver-related diseases. The genetic distribution and identification of HBV genotypes and sub-genotypes represented a challenge to control the spread of infection. To find out molecular prevalence, the present methodology was carried out for the distribution pattern of HBV genotypes and sub-genotypes in north and east regions of India. Methods: A total of 67 HBV DNA positive subjects were studied. At first, the DNA samples for HBV positive cases were screened by Real-time PCR, and then the selected region of HBV polymerase gene was amplified for the sequence analysis to determine genotypes and sub-genotypes. Results: The prevalent genotype found was the genotype D (62.68 %), followed by genotype A (29.85 %), and genotype C (7.46 %). Sub-genotype C1 was identified in the east region only. The frequency of sub-genotype A1 was higher in the north region (n = 13, 30.95 %) followed by east region (n = 7, 28 %). HBV sub-genotype D1 was found to be predominant in 15 (35.71 %) subjects followed by sub-genotypes D2 in 10 (23.80 %) subjects from the north region. HBV sub-genotype D2 was found to be predominant in 8 (32 %) subjects from the east region. Conclusions: In conclusion, the method clearly demonstrates the high prevalence of sub-genotypes D1, D2, A1 in this region. Also, the identification of the sub-genotype C1 in the east region emphasizes the high transmission infection risk and transmission route towards other regions of India.

Keywords: Epidemiology, Phylogenetic analysis, Polymerase chain reaction, Sequencing.

Introduction

Hepatitis B virus (HBV), belongs to family Hepadnaviridae, is a partially double-stranded DNA virus and bears approximately 3.2 kb nucleotides. It has four partially overlapped open reading frames encoding for hepatitis B surface proteins, core peptide, X peptide, and DNA-polymerase enzyme. HBV genome, having sequence heterogeneity, replicates by reverse transcription using a polymerase that lacks proof-reading activity (1). HBV has been circulated 400 million people worldwide (2), with a high prevalence in Asia and Africa. HBV is the most common cause of chronic hepatitis, Hepatocellular carcinoma (HCC), cirrhosis and liver failure (3 and 4). Being a global health problem, some patients affected mildly (2), while other patients developed HCC, cirrhosis and, death from chronic Hepatitis B Virus (5).

Full-length genomes comparisons from different geographical regions classified the HBV into ten genotypes from A to J, based on nucleotide differences by more than 7.5% and further segregate into sub-genotypes, differ from each other by 4 to 7.5% (6). It has been reported that different genotypes and sub-genotypes show different geographical distribution, disease
progression, the response on antiviral treatment, and prognosis, but the mechanisms of different pathogenicity of HBV genotypes are unknown. More than 30 sub-genotypes belonging to HBV genotypes have been determined. Genotypes A, B, C, D, F and I are divided into sub-genotypes, whereas no sub-genotypes have been defined for genotype E, G, and H (7). Genotype D is most common and consist 9 sub-genotypes, where D1 to D3 appear worldwide while D4 to D9 have limited distribution (8). Genotype Ae has been defined in Europe and Aa in Asia along with Africa (9). For the classification of the genotype C, Sub-genotype C1 is found in South-east Asian countries like Thailand, Myanmar, Vietnam, C2 in East Asian countries like Japan, Korea, China (10), C3 in Oceania comprising strains and C4 in Aborigines from Australia (11).

Although HBV viral load is necessary to identify the risk for HCC, cirrhosis, and death in hepatitis infection (5), it is also proposed that viral load is essential, but not sufficient (12). Therefore, other viral markers are required for diagnosis and treatment of the patient. The viral load, circulating genotype and specific risk behavior in a region can help the information available in the country to support the development of specific prevention strategies for the exposed population. Very limited scientific studies are available on HBV genotyping in Indian subcontinents (13). The other challenge for molecular epidemiology in India is to determine the reliable target region to identify genotypes and sub-genotypes with the simple and one experimental process. In this study, we have carried out to find the prevalence and distribution of HBV genotypes with sub-genotypes among the HBV infected subjects in the Indian region. It is also covered the aspects of using simple method to identify molecular epidemiology and distribution pattern of HBV genotypes and sub-genotypes in north and east regions of India.

Materials and Methods

Ethical statement

This study was supported by Kumaun University, Nainital and CORE Diagnostics, Gurugram and approved by its ethics committee. The informed consent form was obtained from all the subjects or patients included.

Study subjects

A total of 83 infected subjects with hepatitis B were enrolled in this study. The study subjects were covering the patients of the northern region (n = 53), from the Delhi (New Delhi = 50), and Uttar Pradesh (Ghaziabad = 3) and the eastern region (n = 30), from the Bengal (Kolkata = 6), Bihar (Patna = 10), Orissa (Cuttack = 2, Ling-raj = 1), the eastern country region; Bangladesh (Dhaka = 9) and Nepal (Kathmandu = 2) that were diagnosed with HBV DNA infection. Northeast states of India were not the part of our study. The subjects that recruited between June 2017 and December 2017 were part of our study. Blood was collected from the subjects for HBV diagnosis and processed in the laboratory. The samples were tested by commercial Real-time PCR (QIAGEN, Germany) to determine the HBV DNA infection.

The mean age of the patients included in this study was 32.02 ± 15.21 years. Plasma was separated from blood and stored at -20°C until the extraction was done. After that, samples were tested for HBV Genotyping or sub-genotyping by using specified primers.

Viral DNA extraction and HBV DNA confirmation

Viral DNA was extracted from 500 μl of plasma samples by QIAamp® DSP virus extraction procedure using silica column-based technology (QIAGEN, Germany). HBV DNA was detected and quantified by artus® HBV RG PCR kit (QIAGEN, Germany) according to the instruction of manufacture on the Real-time PCR system (Rotor-Gene Q, QIAGEN, Germany).

Genotyping with HBV polymerase gene region

Primers used were; forward: 5'- TCGTGG TGGACTTCTCTCAATT-3’ and reverse; 5'- CGTTGACAGACCTTTCGAATC- 3’ for the partial HBV polymerase gene region (14). The composition of 30 μl reaction volumes of PCR master mix was contained 10X PCR buffer, 50
mM MgCl₂, 10 mM each of the four dNTPs, 10 μM of each primer with a final concentration of 0.33 μM and 5U of Taq DNA polymerase. The temperature parameters were; 95°C for 15 minutes, followed to 45 PCR cycles at 95°C for 45 seconds, 56°C for 45 seconds, and 72°C for 45 seconds. Around 730 bp products recovered on agarose gel on amplification with primers. After that, clear and strong bands of PCR positive samples were selected for direct sequencing. The samples that shown weak bands or did not produce satisfactory sequence were excluded from the study.

Sequencing of the amplified product
The composite PCR products were purified using spin column-based purification kit according to the manufacture instruction, and the purified product was measured by Nanodrop spectrophotometer. The Reverse primer was used as a sequencing primer for all the samples. Amplified PCR products were directly sequenced in the ABI 3500xL Genetic analyzer (Applied Biosystem, USA) Instrument, using the Bigdy terminator (Version 3.1) cycle sequencing kit. For the sequencing, thermal cycling, conditions used were 20 seconds on 95°C, 25 seconds on 50°C for 35 cycles, and 60°C for 2 minutes. Data collection and assembly were done by 3500xL Genetic Analyzer data collection and sequencing analysis software (Version 1.0 and 5.4).

Sequence analysis
Sequences were analyzed using KB™ Basecalling (Version 1.4.1.8) sequence analysis software. Sequences received were edited in Chromas software (Version 2.6.4.0) with the comparison of known sequences from the Genbank database, and saved as FASTA file format. Multiple sequence alignments and phylogenetic analysis were done using reference sequences available from Genbank.

Genotyping & Sub-genotyping determination
Obtained sequences were aligned with published sequences from the GenBank database with known HBV genotypes and sub-genotypes (15, 16 and 17). Multiple sequence alignment was performed by using CLC Sequence Viewer, Version 6.1 (CLC, Denmark) software. HBV Genotype and sub-genotype were determined by phylogenetic analysis in CLC Sequence Viewer, Version 6.1 using the neighbor-joining method with a bootstrap analysis of 1000 replicates. This analysis is based on partial reverse transcriptase regions of HBV polymerase sequence. Genotype and sub-genotype of HBV were also determined by the Basic Local Alignment Search Tool (BLASTN) program, available in http://www.ncbi.nlm.nih.gov/ projects with reference to viral nucleotide sequences. That was done by BLAST of the query sequence with the known set of sequence (18).

Accession numbers on submission
After analyzing the 67 HBV polymerase gene sequences, we submitted our sequenced to GenBank, BioSample submission as SUB4043488, and we obtained accessions number from SAMN09237416 to SAMN09237482 for our ascending order sequence.

Statistical analysis
Statistical analysis was done using Microsoft Excel. Demographic Variables were measured as mean ± SD. Comparisons for categorical variables were analyzed using the Fisher’s exact test and the difference for a p-value of <0.05 was considered statistically significant.

The informed consent form was obtained from all the subjects or patients included, and the study approved by ethics committee of CORE diagnostics, Gurugram with permission of Kumaun University Nainital.

Results
The study subjects that included to defined genotypes and sub-genotypes were from north and east regions of the Indian subcontinent. Subjects from the north region identified (n = 42) were of Delhi (New Delhi = 40), and U.P. (Ghaziabad = 2), Subjects from the Bengal (Kolkata = 5), Bihar (Patna = 10), Orissa (Cuttack Prevalence of hepatitis B virus genotypes and sub-genotypes in India
The limit of quantification with the Real-time PCR was the 10.5 IU/ml for the HBV DNA viral load. The HBV DNA levels measured using TaqMan Real-time PCR (QIAGEN, Germany), for the samples used in sequencing study were the ranged from 500 IU/ml to $2 \times 10^7$ IU/ml. When amplified the positive sample with specific primers, PCR product of approximately 730 bp was revealed on agarose gel electrophoresis (Figure 1), of the viral load ranged from 608 IU/ml (limit of detection) to $2 \times 10^7$ IU/ml. Furthermore, 67 samples were successfully sequenced and analyzed.

**HBV genotypes**

Total of three HBV genotypes A, C, and D was found in this population. Of the 67 samples, 42 were detected as genotype D, found to be the predominant circulating genotype (62.68 %). 20 were the genotype A (29.85 %) and 5 were the genotype C (7.46 %). The observed sequences [GenBank: Biosampleaccessions number; SAMN09237416 - SAMN09237482] were aligned with the reference sequences available from GenBank for all known HBV genotypes and determined by constructing a phylogenetic tree. The genotypes of observed sequences were also identified by the genotyping tool (NCBI) with same results.

**Distribution of HBV genotypes within the region**

Distinct patterns of the HBV genotype identified in the study populations from the regions are shown in Table 1. Among the 42 subjects from the north region, HBV genotype D was the predominant genotype, identified in 29 (69.04%) subjects compared to genotypes A identified in 13 (30.95%) subjects ($p < 0.00$). Among the 25 subjects from the east region, HBV genotype D was predominantly identified in 13 (52%) subjects. The other genotypes A and C were identified in 7 (28%) and 5 (20%) subjects, respectively ($p < 0.00$).

**Figure 1**: Agarose gel electrophoresis showing PCR positive samples in the lane S1 to S6 (band size approximately 730 bp) for HBV partial polymerase gene sequence and a negative control in lane S7, followed by ladder 100bp.

**HBV sub-genotypes**

Among the 20 genotype A sequences, all identified as sub-genotype A1 (Figure 2). HBV genotype D sequences were clustered with sub-genotypes D1 (n = 16, 38.09%), D2 (n = 19, 45.23%), D3 (n = 4, 9.52%) and D5 (n = 4, 9.52%) (Figure 3). All 5 HBV genotype C sequences were identified as sub-genotype C1 (Figure 4).

**Distribution of HBV sub-genotypes within the region**

The distribution of HBV sub-genotypes with their frequency are shown in Table 1. Sub-genotype C1 was identified only in the east region. The frequency of sub-genotype A1 was higher in the north region (n = 13, 30.95%) followed by east region (n = 7, 28%). Sub-genotypes D1, D2, D3 and D5 were identified in 15 (35.71%), 10 (23.80%), 2 (4.76%) and 2 (4.76%) subjects from the north region, respectively ($p < 0.00$). HBV sub-genotype D2 was found to be predominant in the east region (n = 8, 32 %) followed by D3, D5 of each with 2 (8%) subjects and D1 with 1 (4 %) subject ($p < 0.00$).
Prevalence of hepatitis B virus genotypes and sub-genotypes in India

Discussion

The most common genotype was genotype D (62.68%) that was present in these regions; supported the observation from other parts in India (3). Another genotype identified was the genotype A (29.85%) in north and east region also reported earlier in India (19 and 20), followed by sub-genotypes and countries of origin. 20 subjects of HBV polymerase gene partial sequences determined in this study indicated by respective keys starting with “SEQ”.

Figure 2: Phylogenetic analysis for the sub-genotypic distribution of HBV genotype A. The tree was formed by taking 21 HBV polymerase gene partial sequences belonging to different sub-genotypes of A1 to A7, indicated by accession numbers, followed by sub-genotypes and countries of origin. 20 subjects of HBV polymerase gene partial sequences determined in this study indicated by respective keys starting with “SEQ”.

Figure 3: Phylogenetic analysis for the sub-genotypic distribution of HBV genotype D. The tree was formed by taking 41 HBV polymerase gene partial sequences belonging to different sub-genotypes of D1 to D9, indicated by accession numbers, followed by sub-genotypes and countries of origin. 42 subjects of HBV polymerase gene partial sequences determined in this study indicated by respective keys starting with “SEQ”.

Prevalence of hepatitis B virus genotypes and sub-genotypes in India
represented the distinct distribution of genotypes particularly in north and east region (22 and 23). Genotype C identified in east region in our study also, where we identified four cases of Dhaka and one case of Kolkata suggests the transmission route from eastern countries (23 and 24). HBV genotype C infection has previously been defined with more active or severe liver disease in Southeast Asia (13). Therefore, need to monitor carefully in north and east region in India as well. Genotype A was also prevalent in earlier studies (19 and 25). HBV genotypes A has less clinical significance compared to genotypes C and D, which is more prone to the development of cirrhosis and HCC (26 and 27), emphasize the major attention to stop the circulation of genotypes in these regions.

On the observation of genomic group C, we found only single sub-genotype C1. For the genomic group A, we found one sub-genotype A1, which is more prevalent in India as well as in the world (28). However, in case of genomic group D, multiple sub-genotypes D1, D2, D3 and D5 were found (Table 1) (3). In the phylogenetic tree also, the majority of the HBV sub-genotypes branched in the D1 group (38.09%, 16/42) and in the D2 group (42.85%, 18/42). However, 9.52% (4/43) of each sub-genotype branched in D3 and D5 groups (Figure 3).

To eliminate the confusion of reaction failure or to use another internal control, sequencing reactions were preceded with the same extracted materials that were used for HBV DNA viral load identification. These positive Real-time PCR results were again confirmed by targeting the HBV polymerase gene region, and recovery of amplified product on the agarose gel (Figure 1). The process has been utilized by using one set of primer, to identify and analyze genotypes as well as sub-genotypes in a single experimental run. Therefore, it is convenient and effective method to perform the experiment and analyze the results. This is done by aligning and creating phylogenetic tree of the different genotyping sequences with known sub-genotype reference sequence (Figure 2, 3, 4). In our knowledge, the above applied methodology of HBV DNA sequencing has not been demonstrated within Indian subcontinent and may be a useful target for the disease management of viral hepatitis patients.

A limitation of this study was the lack of information about the previous diagnostic clinical history of the patient, whether they were carriers of HBsAg, chronic Hepatitis B or sufferings decompensate cirrhosis, and their life-style related to viral transmission.

Jagdish et al
Conclusion
In conclusion, we have demonstrated the prevalence and distribution of HBV genotypes and sub-genotypes in the north and east regions of India. The method utilized of genotyping and sub-genotyping for HBV infection is useful tools to identify the epidemiology of HBV infection and to understand the clinical significance for the associations of disease progression. The high prevalence of sub-genotypes D1, D2, A1 and other sub-genotypes clearly suggest the transmission risk to the population and circulation subjects from these regions. Also, the presence of a sub-genotype C1 in the east region and neighbor countries can be the higher transmission risk to the east region and could be the transmission route towards other regions of India. However, further analysis is required with a large number of sample volumes to define and distinguish the sub-genotypes by using complete genome sequence.

Acknowledgments
We thank the staffs of Chromous Biotech, Bengaluru, and Lifetech Services Laboratory, Gurugram, Invitrogen Bioservices, India for assistance in the direct sequencing studies.

Funding
No funding or financial support declares.

Conflicts of interest
No conflict of interest.

References

Prevalence of hepatitis B virus genotypes and sub-genotypes in India


Prevalence of hepatitis B virus genotypes and sub-genotypes in India


