Analysis of Complete and Partial Genome Sequences of Hepatitis B Virus and Determination of its Genotypes and Sub-Genotypes from Pakistan

Majid Mahmood,1,6,*, Abida Raza,2 Muhammad Asim Anwar,3 Mazhar Qayyum,1 Nasib Zaman,4,7 Azra Khanum4,5 and Mirza Azhar Beg1

1Department of Zoology, PMAS Arid Agriculture University, Rawalpindi, Pakistan
2Diagnostic Department, Nuclear Medicine, Oncology and Radiotherapy Institute (NORI), Islamabad, Pakistan
3Department of General Medicine, PAEC General Hospital, Islamabad, Pakistan
4Department of Biochemistry, PMAS Arid Agriculture University, Rawalpindi, Pakistan
5Barani Institute of Management Sciences (BIMS), Rawalpindi
6Department of Zoology, University of the Poonch, Rawalakot, Azad Kashmir
7Center for Biotechnology & Microbiology, University of Swat, Swat, Pakistan.

ABSTRACT

Hepatitis B virus has been classified into eight genotypes from A to H which have variety in geographical distribution. HBV/D is reported to be the most prevalent genotype in Pakistan. Sub genotypes D1 and D3 were reported in a study from Karachi. Present study was designed to investigate the genotypes and sub-genotypes of HBV with sequencing of S gene fragment as well as complete viral genome of samples from Islamabad, Pakistan. A total of 22 HBV sample were genotyped by type specific primer PCR method. Complete viral genome of 4 samples and the S gene fragment of all samples was subjected to direct sequencing. Genotypes and sub-genotypes were confirmed by phylogenetic analyses of the sequences. Employing type specific primer PCR method, 20 of the sequenced samples were identified as genotype D and not genotype A or mixed, while complete genome sequences of 4 were identified as sub-genotype D1. Thus this study provides evidence for prevalence of D genotype more specifically sub-genotype D1 in Islamabad. Furthermore, it is observed that sequence based results are more reliable as compared to type specific PCR based method.

INTRODUCTION

Hepatitis B virus genome is greatly variable having diverge genetics in different regions of the world. On the base of at least 8% divergence in its genome, HBV is segregated into eight different genotypes named from A to H (Okamoto et al., 1988; Naumann et al., 1993; Norder et al., 1994; Stuyver et al., 2000; Arouz-Ruiz et al., 2002; Huy et al., 2008; Olinger et al., 2008; Kanako et al., 2009). The genotypes have specific pattern of distribution in geographic regions and ethnic groups of the world (Stuyver et al., 2000; Kidd-Ljunggren et al., 1995; Bowyer et al., 1997; Datta, 2008; Bonino et al., 2010; Nakano et al., 2001).

Currently, HBV genotyping is not used as a diagnostic test before the treatment of chronic HBV patients in Pakistan and the phylogenetic data of HBV strains in Pakistan is not sufficient. According to the available studies, genotype D is predominant in Pakistan along with a little proportion of genotype A (Hanif et al., 2013; Alam et al., 2007; Noorali et al., 2008; Baig et al., 2009; Ali et al., 2011). Mix infection with genotypes A and D is also highly prevalent in Pakistan (Hanif et al., 2013; Alam et al., 2007; Baig et al., 2009; Awan et al., 2010; Abbas et al., 2006). Major HBV genotypes are also classified in to sub-genotypes on the base of divergence between 4-8% in the complete viral genome (Yousif and Kramvis, 2013). Genotype D is further classified into six sub-genotypes which are confirmed so far (Yousif and Kramvis, 2013). Sub-genotypes D1 and D3 of genotype D have been reported from 11 and 2 samples respectively by the only study from Karachi, a city of Pakistan (Baig et al., 2009).

Clinical significance of HBV genotypes has been studied in various regions of the world. The genotypes affect disease progression, patients’ viral load, HBsAg sero-clearance, HBeAg sero-conversion and response to antiviral therapy in many studies (Norder et al., 2004; Cao, 2009; Erhardt et al., 2005; Bonino et al., 2007; Chan et al., 2009). Along with genotypes of HBV, sub-
genotypes are also significant clinically and, though still not well established, have their impact on course and treatment of the infection (Yin et al., 2008). HBV genotypes and sub-genotypes are needed to be explored, especially in most of the northern and northwestern regions of Pakistan. The present study was designed to identify the sub-genotypes of HBV circulating in the city of Islamabad, Pakistan by sequencing of partial as well as full length HBV genome.

MATERIALS AND METHODS

Samples
A total of 22 samples were collected from patients visiting Nuclear Medicine, Oncology and Radiotherapy Institute (NORI), Islamabad, during the period of January 2012 to April 2012. The samples were the aliquots of plasma or serum originally collected for routine diagnostic purposes. The sera were separated and stored below -70°C until further analysis. All the patients signed an informed consent and the study was approved by institutional ethics committee.

Genotyping using type specific primer method
The DNA was extracted using QIAamp® DNA blood mini kit (Qiagen, Hilden, Germany) from 200µl serum following manufacturer’s protocol. DNA was eluted in 50 µl PCR grade water and stored below -70°C until further analysis.

Type specific primer- PCR method (Naito et al., 2001) was used for HBV genotyping at Nuclear Medicine, Oncology and Radiotherapy Institute (NORI), Islamabad, with a few modifications in thermal profile and reagents as described earlier (Hanif et al., 2013).

Genotyping by direct sequencing of S gene
The samples were genotyped by direct sequencing of an S gene fragment of 685 nucleotides. QIAamp MinElute Virus Spin Kit (Qiagen) was used to isolate HBV DNA following manufacturer’s protocol. Five µl of extracted HBV DNA was amplified in a PCR mix including Tris HCL buffer, 2 mM Magnesium Chloride, 0.1 mM dNTPs, 2 units Dream Taq polymerase (Fermentas) and 0.25 µM each of the primers, LAMes and POL4as (Table I). The PCR thermal cycling profile was: 3 minutes at 94°C then 35 cycles including 30 sec at 94°C, 30 sec at 55°C and, 30 sec at 72°C, and then 7 min at 72°C. Negative samples after first round PCR were amplified in semi-nested PCR using internal HBV reverse primer POL181as (Table I) and a thermal profile like the first round but repeated for 20 cycles instead of 35. After confirming by gel electrophoresis, the products were purified by ethanol precipitation protocol (Beckman Coulter, USA) and subjected to sequencing by chain termination method using commercially available Kit (Beckman Coulter, USA). Briefly, 2 µl of purified DNA was mixed with 8 µl of DTCS Quick Start Master Mix, 4 µl of 1.6 µM primer (LAMes and POL181as) and 6 µl of water. The thermal profile used was; 30 cycles of 20 sec at 96°C; 20 sec at 50°C & 4 min at 60°C. The instrument CEQ 8000 XL Analysis System (Beckman Coulter, USA) was used for sequencing.

Complete genome sequencing
Complete HBV genome was sequenced for 4 samples. The genome was sequenced in 7 different overlapping fragments. Each fragment was sequenced using separate pair of primers according to the method described above. The sequences of primers are given in Table I. All the sequencing and phylogenetic analysis was performed at Hepatology Laboratories, University Hospital of Pisa, Italy.

Analysis of sequences
Initially sequences were observed and corrected manually in Chromas-Lite, Version 2.1.0. Initial alignment, editing and joining of different genome fragments was performed by Bio-edit and M ultalin online tools. The sequences were finally analyzed with MEGA version 4.0 to obtain the phylogenetic trees using maximum composite likelihood with 500 bootstraps replication in neighbor joining method. Number of base differences per sequence and number of base differences per site were calculated by the distance analyses conducted in MEGA4 using complete deletion option in which all the positions containing gaps and missing data were eliminated from the dataset. There were a total of 3092 positions in the final distance analyses dataset.

RESULTS

Genotyping
Out of the total 22 samples, genotype D was identified in 14 (63.6%), while 8 (36.4%) samples were found to be infected with a mixture of genotypes A and D (A+D) using type specific primer PCR method.

On direct sequencing genotype D was detected in 20 (91%) samples out of the total 22 samples, while sequencing was not successful for 2 samples. Neither of the samples was detected with genotype A or the combination of A+D. In phylogenetic analysis, all the 20 sequences were clustered with the sequences of genotype D and none of them was clustered with sequences of other genotypes (Fig. 1).
Fig. 1. The phylogenetic tree of 20 partial S gene sequences with reference sequences of all HBV genotypes and sub-genotypes confirming genotype D.

Fig. 2. The phylogenetic tree of 4 complete genome sequences with reference sequences of all HBV genotypes and sub-genotypes confirming sub-genotype D1.
The present study was designed mainly to explore genotypes/sub-genotypes of hepatitis B virus in Islamabad, Pakistan and then to confirm the mix genotype infection of genotype B and D with direct sequencing of infection of genotype B and D with direct sequencing of S gene fragment as well as complete genome of HBV.

**DISCUSSION**

The present study was designed mainly to explore genotypes/sub-genotypes of hepatitis B virus in Islamabad, Pakistan and then to confirm the mix genotype infection of genotype B and D with direct sequencing of an S gene fragment as well as complete genome of HBV.
When genotyped by type specific primer PCR method, the method used commonly to genotype HBV in Pakistan, we found genotype D alone from 63.6% of the samples and mix infection with genotypes A and D simultaneously from 34.4% of the samples. However, with direct sequencing method, we did not identify any sample with dual infection and all the successfully sequenced samples had genotype D sequence. The primer PCR method amplifies different fragments of different genotypes and the primers used are specific to the genotypes and they did not amplify the positive controls of non-specific genotypes when tested for confirmation. Moreover, in many types of confirmatory tests including different types of positive and negative controls, the mix genotype infection was confirmed by the type specific primer PCR method. The mix infection with the combination A+D was also reported by some other studies from Pakistan (Hanif et al., 2013; Alam et al., 2007; Baig et al., 2009; Awan et al., 2010; Abbas et al., 2006) which confirm the presence of this type of infection. The reason may be the sensitivity difference of both the methods used as it is possible that genotype A which was not detected in any of the sample by direct sequencing method, has very less number of copies in the sample as compared to genotype A copies and was, consequently, not detected by the sequencer. The genotype A may have a low replication capacity or at least in the presence of another genotype such as D. This may be because of some competition for the replication machinery inside the cell or for new cell infection. All this is required to be studied in a larger number of samples infected with both genotypes A and D simultaneously.

Sub-genotypes D1, D2 and D4 were recorded in 9, 4 and 7 of our samples respectively by direct sequencing of S gene fragment while D1 was identified from all the four samples subjected to full length genome sequencing. Sub-genotypes D1 and D3 are already reported from Pakistan (Baig et al., 2009). Sub-genotypes D1, D2, D3 and D5 are reported by more than one studies from Eastern India (Banerjee et al., 2006; Datta et al., 2008) while D1 and D2 are also reported from Turkey (Bozdayi et al., 2005). So the presence of D1 and D2 genotypes in Pakistan, where genotype D is the predominant one, is quite likely. The presence of sub-genotype D4 which was identified in 7 out of 20 samples, however is surprising as this genotype was neither previously reported from Pakistan nor from any of the neighboring countries but it is mostly present in Australia, Oceania, Western Africa and South America (Norder et al., 2004; Kwange et al., 2013; Santos et al., 2010). So, a study is needed for the confirmation of this sub-genotype from Pakistani isolates of HBV. This study concludes that genotype D and sub-genotype D1 of hepatitis B virus are dominant in Islamabad. Moreover, mixed HBV genotypes, specifically the mixtures of A and D genotypes detected by type specific PCR method has not been confirmed by direct sequencing method strengthening the reliability of direct sequencing method over type specific PCR method for HBV genotyping.

ACKNOWLEDGMENT

Funding for the study was provided partially by Higher Education Commission (HEC) of Pakistan in the form of Indigenous PhD Fellowship and International Research Support Program (IRISIP). The authors are thankful to Prof. Ferruccio Bonino who provided opportunity to work in his laboratory at University hospital of Pisa, Italy for this study and Dr. Francesco Moriconi who revised the manuscript.

Declaration of conflict of interest

The study is a part of PhD thesis of Majid Mahmood. There is no other conflict of interest among the authors.

REFERENCES


