Phylogenetics and Evolutionary Association of Hepatitis B Virus Isolated from Pakistan

Ali Raza Awan,1,3 Masroor Ellahi Babar,1 Ahmad Ali,1* Zara Awan2, Zafar Saleem and Yasir Zahoor3
1 Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Outfall Road, Lahore, Pakistan
2 Agha Khan University Medical College, Stadium Road, Karachi 74800, Pakistan
3 Centre of Excellence in Molecular Biology, University of Punjab, 87, West Canal Bank Road, Thohkar Niaz Baig, Lahore

Abstract.- Genetic variations of Hepatitis B Virus (HBV) are closely associated with viral pathogenesis. The generation of variants of Hepatitis B virus (HBV) with altered functional properties and increased pathogenesis are the result of mutational events that can affect the disease patterns and response to antiviral treatments. Due to high rate of mutational changes in viral genomic structure the resulting variables become resistant to antiviral drugs therefore making it imperative to develop therapeutic alternatives by identifying rapidly occurring variants. This study describes genetic variability of Pakistani HBV isolates based upon DNA sequences of cloned PreS1/PreS2/S region or PreS/S Open Reading Frame (ORF). HBV DNA was isolated from blood samples of 10 unrelated patients and PreS/S ORF of HBV was amplified through PCR from these isolates. The DNA fragments were cloned and sequenced using standard methodologies. The surface genes from HBV isolates showed 96-99% homology with reported HBV genotype D but Pakistani isolates demonstrated sequence novelties never reported elsewhere. Phylogenetic analysis of the sequences revealed that these isolates clustered within the genotype D group reported in different regions of the world. This study describes the phylogenetic and evolutionary characterization of HBV isolated from Pakistan based on DNA sequence of complete PreS/S ORF.

Keywords: HBV, hepatitis, genotype, evolution, phylogenetics, Pakistan

INTRODUCTION

Hepatitis B infection happens to be a major health concern in Pakistani urban and semi-urban populations. More than 350 million people suffer from chronic Hepatitis B Virus (HBV) worldwide and it is estimated that more than two billion people are at risk (McMahon, 2005). Carrier variability rate for Hepatitis B infection is estimated to be 0.1% to 20% throughout the world (Saleh et al., 2005). An estimated 8% South Asian population suffers from this infection at least once in their lifetime with HBV (Lindh et al., 1999).

Eight genotypes of HBV have been reported so far throughout the world having distinct geographical distribution patterns. Genotype A has been reported in Northwest Europe, North America, Philippines (Norder et al., 1993b; Kidd-Ljunggren et al., 1995), Hong Kong (Lok et al., 1994) and similarly Genotype B and C are predominantly found in Southeast Asia (Okamoto et al., 1988; Kidd-Ljunggren et al., 1995; Theamboonlers et al., 1999). Genotype D is however the most extensively distributed and has been reported throughout the world. It is mostly found in the region stretching from South Europe and North Africa (Norder et al., 1993b; Borchani-Chabchoub et al., 2000) to India, in the West and South Africa (Bowyer et al., 1997). Genotype E was first described as a subset of D (Kidd-Ljunggren et al., 1995) but this subgroup was later classified as an independent genotype mostly found in West and South Africa (Norder et al., 1993a, b, 1994). Genotype F is the most divergent of all and is found in Central and South America (Norder et al., 1993a; Arauz-Ruiz et al., 1997a, b; Blitz et al., 1998; Mbayed et al., 1998; Nakano et al., 2001). The genotype G has only been reported in USA and France (Stuyver et al., 2000). Arauz-Ruiz et al. (2002) reported a new genotype H, having profound similarity to genotype F, and known to be an Amerindian genotype. This genotype is probably a split fragment of F genotype within the new world by early division of the progenitor HBV strains of the first settlers (Arauz-
Ruiz et al., 2002). The criteria for grouping the HBV isolates into different genotypes comprises of at least 92% homology between sequences of the S-gene or a minimum inter-genotypic score of 4.1% (Mizkomi et al., 1999; Bowyer and Sim, 2000; Stuyer et al., 2000). It has also been shown that genetic analysis on the basis of the S-gene are comparable to genotyping complete HBV genomes (Norder et al., 1993; Ohba et al., 1995; Mizkomi et al., 1999; Bowyer and Sim, 2000; Stuyer et al., 2000) On the basis of variation analysis, homology of surface genes and full genome sequences, a specific HBV genotype is defined as a subtype when there is at least 92% sequence similarity between genotypes (Magnius and Norder, 1995). This study was aimed determining the genotypic grouping of Pakistani HBV isolates and the variation profile reported genotypes in other parts of the world. The phylogenetic analysis revealed evolutionary association of Pakistani isolates with reported genotype D of HBV.

MATERIALS AND METHODS

Vector and strains
The T-A cloning vector containing 3’ terminal thymidine at both ends (PCR2.1 vector) and E. coli Top10F’ strain for plasmid manipulation was purchased from Invitrogen Co. USA.

Collection of blood samples and DNA extraction
Peripheral blood samples were obtained from 25 patients having positive surface antigen markers from different hospitals in the Punjab province of Pakistan. These patients were screened for HBsAg, anti-HBs, HBeAg, anti-HBe, anti-HCV and anti-HIV; patients with anti-HCV co-infection were excluded. Viral DNA from patients was isolated through proteinase K digestion method (Persing et al., 1993).

Primer selection
Forward and Reverse Primers were designed using the most conserved flanking region of PreS/S ORF previously reported HBV sequence (accession no. NC_003977). Web based freeware www.primer3.com was utilized to design accurate primers with optimal GC content and melting temperatures.

PCR amplification and gene cloning
The PreS/S ORF of HBV was amplified through PCR from acquired isolates using primers (sense 5’ TATTCTTGGGAACAAGAG 3’ and antisense 5’ GCAGCAAAGGCCAAAAG 3’). PCR was optimized using 50ng of template, 10 picomole of each primer, 2 units of Taq polymerase, 0.2mM of each dNTPs and different thermo cycling programs. These DNA fragments were cloned into a T-A cloning vector and confirmed through PCR and restriction digestion. To control for misincorporation of nucleotides by Taq polymerase, two independent clones of each viral isolates were selected for further analyses. Using Big Dye Terminator Cycle Sequencing Ready Reaction Kit on ABI-3100 DNA analyzer, cloned DNA fragments were sequenced.

Genetic variability analysis
A total of 43 reference sequences were retrieved from GeneBank and were used for genetic variability analysis of Pakistani HBV isolates. The accession numbers with their respective country of origin are; Genotype A; AB194951 from Cameroon, AB014370 from Japan, EU410082 from Philippines, AF297621 from South Africa, AB222707 from Uzbekistan, Genotype B; M54923 from Indonesia, AB073838 from Japan, AF121243 from Sweden, AY167098 from Taiwan, AY167098 from the UK, Genotype C; EU439009 from China, AB014393 from Japan, AB105172 from Hawaii, X97850 from Polynesia, Genotype D; AF280817 from China, AB03558 from Japan, EF103276 from India, AY741798 and AY741797 from Iran, DQ991753 from Ireland, X65257 from Italy, AB263407 from Mongolia, AB033559 from Papua New Guinea, Z35716 from Poland, AF121240 from Sweden, X02496 from Switzerland, AY661753 from Turkey, AF121239 from Vietnam, X80924 from the UK, AB222709 from Uzbekistan, Genotype E; AM494832 from Central African Republic, AB106564 from Ghana, X75664 from Senegal, X75657 from West Africa, Genotype F; X69798 from Brazil, X75658 from France, AB036911 from
Venezuela, Genotype G; AF160501 from Belgium, AF405706 from Germany, AB056513 from the USA, Genotype H; AY090454 from Sweden. The HBV PreS/S ORF (1170bp) and reference sequences were aligned with ClustalW freeware by using BioEdit version 7.0.5 (Hall, 1999). Genetic distance was calculated using Kimura two-parameter matrix (Saitou et al., 1987). Phylogenetic trees were constructed by the neighbor-joining (NJ) method (Kimura, 1980).

RESULTS AND DISCUSSION

DNA sequences of PreS/S ORF (1170bp) of ten HBV isolates were submitted to NCBI GeneBank database under accession numbers from FJ670505 to FJ67014. After alignment with 43 HBV genotypes retrieved from the GeneBank the homology analysis showed that the local isolates had 96% to 99% similarity with genotype D of HBV (Fig. 1). The Phylogeny of isolates was analyzed using DNA sequences of isolates and reference HBV genotypes to further confirm that Pakistani HBV isolates belonged to genotype D (Fig. 1) All isolates showed arginine residue at amino acid position 122 determining subgroup “y” and a lysine residue at AA position 160 determining subgroup “w”, collectively confirming the serotype of the isolates as “ayw”. Hepatitis B specifically infects hepatic cells of the members of hominoidae including humans. Vast data on HBV variability has now been gathered, however, the topical issues are geographical genetic variation that affect HBV mode of infection, disease pattern, immunization, prophylaxes and treatment. HBV changes itself due to mutations in its genetic material in response to environmental stresses (Bowyer and Sim, 2000). HBV genome has been shown to change at an exchange rate of 0.1 nucleotide per year (Okamoto et al., 1988). Heterogeneity among globally common HBV strains is 104 times greater than other DNA viruses which is due to the fact that members of hepadnaviridae family replicate through an RNA intermediate and RNA reverse transcriptase is known to have a high error rate (Hourious et al., 2000). The diversity of HBV is shown through its

Fig. 1. Phylogenetic tree constructed using Kimura two-parameter matrix and the neighbor-joining method. Pakistani sequences are named from PKHB2-PKHB20. The letters A-H shows the genotypes.
different genotypes and serological subtypes (Ma et al., 2005).

In this study, the complete HBV surface ORFs were cloned, sequenced and the isolates analyzed for phylogeny. The results indicate there was no genotypic divergence amongst the isolates clustering in genotype D with 96-99% homology. Previously reported HBV genotypes A, B and C are determined through PCR based genotyping not by sequencing DNA of isolates (Idrees et al., 2004). In a recent study, genotype D was reported from Pakistan based on 967 bp of the HBV surface ORF (Baig et al., 2008), however, the present study includes complete ORF sequence of PreS1, PreS2 and S genes (1170 bp) for genotypic characterization. The findings of Alam et al. (2007), substantiated that genotype D is the most prevalent one in Pakistan. Genotypic studies on HBV in neighboring countries also confirmed the dominance of genotype D in the region. It has been reported to be the single detectable genotype in Iran and Mediterranean regions (Alavian et al., 2003; Tahan et al., 2003; Bozdayi et al., 2005; Leblebicioglu and Eroglu, 2004; Yalcin et al., 2004; Amini-Bavil-Olyaei et al., 2005), whereas genotypes A and D were reported in India with D as a dominant genotype in chronic liver patients (Kar et al., 2007).

The prevalence of D Genotype in the region may be archeological and anthropological. The ancestors of Caucasians firstly colonized North of the Caspian Sea migrating in three directions thereafter: one group moving to Europe, second to India and the last one moving towards south of Iran (Jazayeri and Carman, 2009). With a common line of descent, their ancestors may have carried the same HBV genome, the genotype now being identified as D. The assortment of HBV genotypes A-H in different parts of the world is likely due to immune pressure (Jazayeri and Carman, 2009). The Phylogram depicted in Figure 1 supports the hypothesis that ancient HBV genomes belonged to the genetic structure closer to that of genotype D (Jazayeri and Carman, 2009). This study also supports the hypothesis that genotype H is a derivative of genotype F due to H clustering closely with F. In the Phylogram (Fig. 1), the genotypes A, B and C are also closely clustered with each another (Arauz-Ruiz et al., 2007).

REFERENCES


(Received 3 December 2009, revised 6 May 2010)