ORIGINAL ARTICLE

Phylogenetic Analysis of Hepatitis B Virus in Pakistan

Saeeda Baig[1], Anwar Ali Siddiqui[2], Runu Chakravarty[3], Tariq Moatter[4], Tayyab Unnissa[5] and Nazr-ul-Hasnain[1]

ABSTRACT

Objective: To identify the distribution pattern of Hepatitis B Virus (HBV) genotype in a group of patients and to study its phylogenetic divergence.

Study Design: An observational study.

Place and Duration of Study: The clinics of Gastroenterology Unit, Ziauddin University, from January to December 2006.

Methodology: Two hundred and one HBV infected patients were genotyped for this study. All HbsAg positive individuals, either healthy carriers or suffering from conditions such as acute or chronic hepatitis, cirrhosis and hepatocellular carcinoma were included. Hepatitis B patients co-infected with other hepatic viruses were excluded. Hepatitis B virus DNA was extracted from serum, and subjected to a nested PCR, using the primers type-specific for genotype detection. Phylogenetic analysis was performed in the pre-S1 through S genes of HBV. The divergence was studied through 15 sequences of 967bp submitted to the DBJ/EMBL/GenBank databases accessible under accession number EF584640 through EF584654.

Results: Out of 201 patients tested, 156 were males and 45 were females. Genotype D was the predominant type found in 128 (64%) patients followed by A in 47 (23%) and mixed A/D in 26 (13%). Phylogenetic analysis confirmed the dominance of genotype D and subtype *ayw2*.

Conclusion: There was dominance of genotype D subtype *ayw2*. It had a close resemblance with HBV strains that circulate in Iran, India and Japan.

Key words: *Hepatitis B Virus. Genotyping. Phylogenetic analysis.*

INTRODUCTION

The study of Hepatitis B Virus (HBV) with its evolutionary behaviour remains completely unclear. The genetic diversity of the genes that encode surface antigen of hepatitis B virus is the basis of identification and classification of different genotypes. The eight genotypes (A-H) of hepatitis B virus and several subgenotypes are based on > 8% divergence over the entire genomic sequence.[1] This diversity is partly due to virus/host interactions and partly due to parallel evolution in geographically distinct areas.[2] Alternatively, HBV strains can be classified serologically on the basis of antibody to surface antigen (subtypes). These subtypes appear in more than one genotype. Therefore, structural and functional differences between HBV genotypes are the main reason behind the severity, complications, treatment and possibly vaccination against the virus. Much remains to be learnt about the full clinical implications of genotypes.[3] The phylogenetic analysis can be utilized to elucidate the history and the origin of HBV. It has become a reliable tool to study the divergence in the HBV genome scattered geographically around the world.[4]

Early studies from 1978 and 1980 have shown that *ayw* and *ayw2* are the most common HBV subtypes in Pakistan.[5,6] In concordance with these results genotype D has been shown to predominate in regions of Pakistan, coexisting with genotype A and a mixture of both genotypes, which are found in lesser percentage.[7] But currently, there is no data available on phylogenetic relatedness of the virus endemic in the country. The present study was aimed to determine the pattern of HBV genotypes, and the phylogenetic relationship of the HBV strains prevalent in a group of Pakistani patients and their comparison with the geographically close countries.

METHODOLOGY

Patients attending Ziauddin University Hospital (ZUH) from January to December 2006, were randomly selected after taking informed consent. Prior to this approval of ZU ethical review committee was obtained. All HbsAg positive individuals, either healthy carriers or suffering from conditions such as acute or chronic

[1] *Department of Biochemistry, Ziauddin Medical University, Karachi.*
[2] *Department of Biological and Biomedical Sciences, The Aga Khan University, Karachi.*
[3] *ICMR Virus Unit, ID and BG Hospital, Campus, Kolkata, West Bengal, India.*
[4] *Department of Molecular Pathology, The Aga Khan University, Karachi.*
[5] *Research Associate, The Aga Khan University, Karachi.*

Correspondence: *Dr. Saeeda Baig, Associate Professor, Department of Biochemistry, Ziauddin Medical University, Clifton, Karachi.*
E-mail: baigsaeeda@yahoo.com

Received February 16, 2008; accepted October 18, 2008.

hepatitis, cirrhosis and hepatocellular carcinoma were included. Hepatitis B patients co-infected with other hepatic viral markers such as hepatitis A, C, D or E were excluded. Ten ml blood was drawn from 250 patients, and serum was separated. All sera were stored in aliquots of 200µl each at -70°C. DNA was extracted from serum using a DNA extraction kit (QIA AMP DNA mini kit 250 reactions Cat # 51306). Fifty micro liter of extracted DNA was concentrated by using DNA speed vac. Ten µl of DNA was used for PCR. For PCR amplification, a modified version of nested PCR, developed by Naito *et al.*[8], was followed. The HBV genome was amplified by sequence specific PCR using the universal primers (P1 and S1-2) for the outer primers, followed by two different mixtures containing type-specific inner primers. The PCR primers were supplied by Gene Link USA. The step-one PCR primers (outer primer pairs) and step-two PCR primers (inner primer pairs) are designed on the basis of the conserved nature of nucleotide sequences in regions of the pre-S1 through S genes, irrespective of the HBV genotypes. P1 (sense) and S1-2 (antisense) were universal outer primers (for 1,065bp product). B2 was used as the inner primer (sense) with a combination called mix A for genotypes A, B, and C. Mix B consisted of antisense primer B2R with a combination of Mix B sense primers for genotypes D, E, and F.

Sequencing of HBV/D isolates was done from the nucleotide sequences of 15 HBV/D isolates recovered from patients with HBV-associated liver disease. The DNA fragments of 1065bp were amplified by PCR, with universal primers P1 and S1-2. Amplification was performed in a 96-well cycler and the PCR products were run on electrophoresis in 1.0% (wt/vol) agarose, stained with ethidium bromide, and observed under the UV light. Standard precautions were taken for avoiding contamination during PCR. The PCR product of 1065bp, obtained with the universal primers P1 and S1-2 were bidirectionally sequenced, after purification on Wizard PCR Preps DNA purification resin. Internal primers to the amplified fragments were developed to prime the sequencing reaction. The sequencing was done using ABI's BigDye terminator cycle sequencing kit and 310 automated fluorescent genetic analyzer.

Chromatograms of all the 15 sequences were studied closely for any polymorphism, overlapping, insertion or deletion and a 967bp sequence was finalized for the study. Phylogenetic analysis was done through BLAST (Basic Local Alignment Search Tool) searches and 150 sequences of HBV isolates from around the world were selected. Fifty eight HBV sequences of the corresponding region were finalized, representing each of the eight genotypes A-H to make a phylogenetic tree. These fifty eight sequences were obtained from the international DNA database (DDBJ/EMBL/GenBank) by random selection from 150 sequences, originating in various countries (Iran, Italy, Japan, Egypt, USA, Latvia, Germany, India, Brazil, S. Africa, Philippines, Bolivia, etc.). The best and the high scoring matches with our sequences were aligned with the CLUSTAL W software program,[9] and the alignment was confirmed by visual inspection.

Phylogenetic analysis of a 967 nucleotide fragment of the S gene employing MEGA, version 2.0,[10] and PHYLIP, version 3.54c,[11] was the basis for HBV genotyping using Kimura's 2-parameter algorithms with the neighbour-joining method. The reliability of different phylogenetic groupings was evaluated using the bootstrap test (1000 bootstrap replications). As almost identical groupings were observed with these tree-building programs, only the MEGA-based tree is presented (Figure 1).

The nucleotide sequences data reported in this paper have been submitted to the DBJ/EMBL/GenBank databases under accession number EF584640 through EF584654. These sequences can be retrieved from http://www.ncbi.nlm.nih.gov/blast/Blast.cgi, EMBL in Europe and the DNA Data Bank of Japan.

Amino Acid Sequence alignment was done for serotype. ExPASy Translate tool, which can be retrieved at http://us.expasy.org/, was used to translate the nucleotides (DNA) sequence of S protein into amino acid sequence. All amino acids were assembled into a figure for comparison (Figure 2).

Statistical analysis was done by using percentages for qualitative variables.

RESULTS

A total of 201 serum samples were genotyped with repeating samples twice and sometimes thrice for confirmation. There were 156 (78%) males and 45 (22%) females with a ratio of 3.5:1. Genotype D with 128 patients (64%) was found to be the most prevalent amongst our patients, rest of the genotypes A, AD mix, were distributed as 23%, and 13% respectively (Table I).

Table I: Genotype distribution in the population studied.

Genotype	Male 156 (78%)		Female 45 (22%)		Total 201	
	Number	Percent	Number	Percent	Number	Percent
A	38	24	9	20	47	23
D	98	63	30	67	128	64
A/D	20	13	6	13	26	13

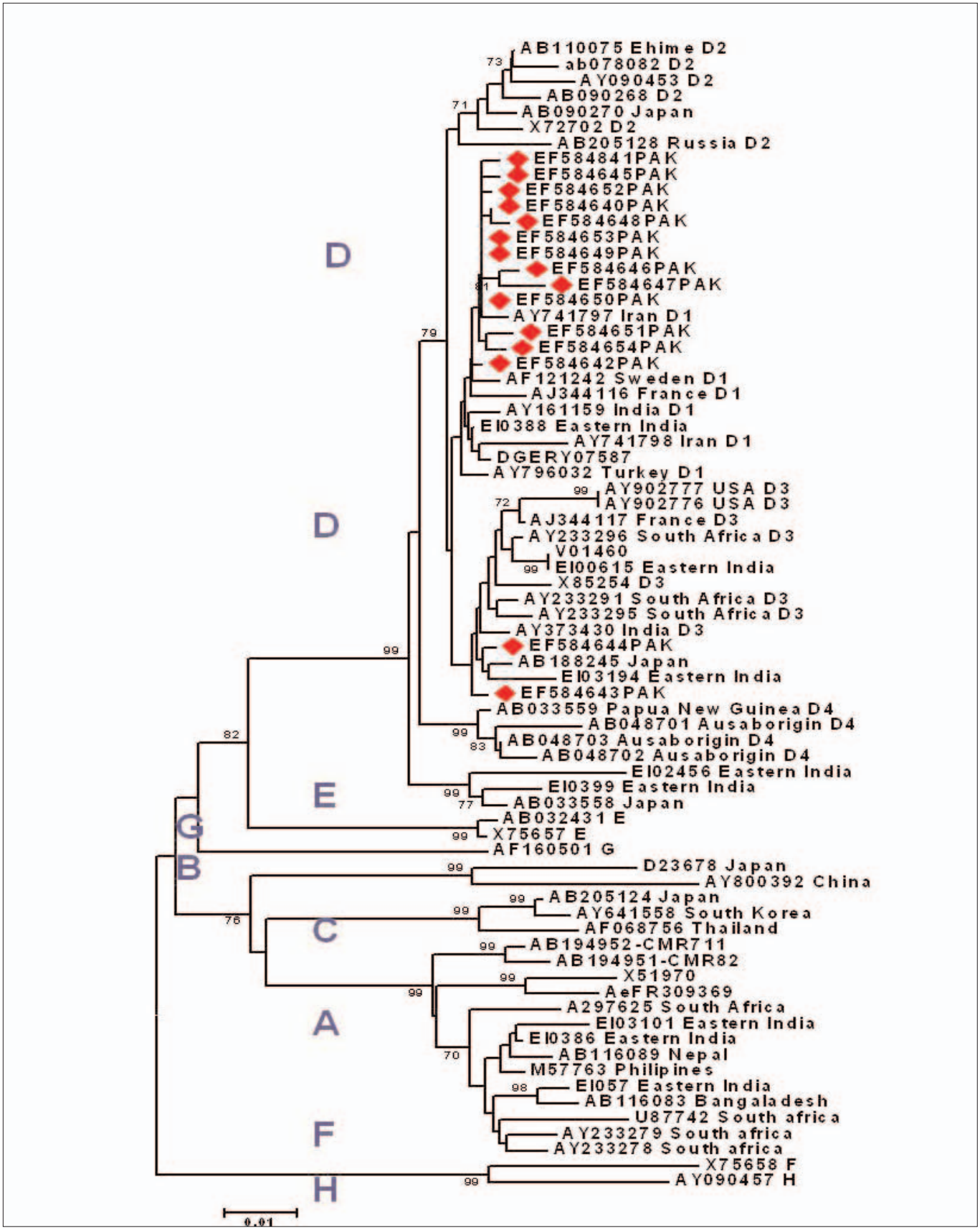

Figure 1: Genetic relatedness of genotype D to other genotypes. DIST= percentage divergence/100). The sequences of this study are highlighted in red. The reliability of different phylogenetic groupings was evaluated using the bootstrap test (1000 bootstrap replications) given as numbers at the nodes. The bar showing the value 0.001 gives genetic distance of the sequence analysed.

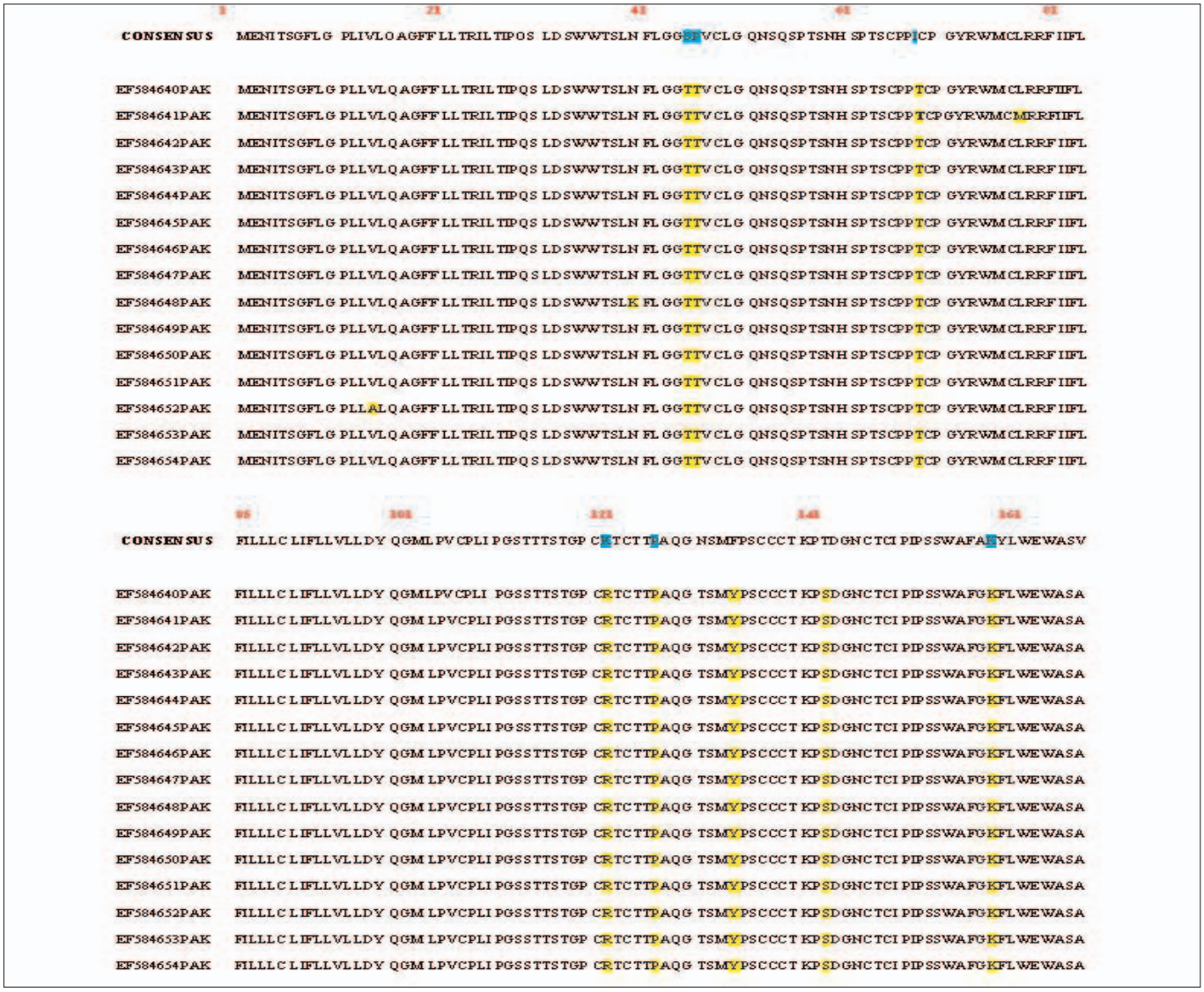

Figure 2: The amino acid residues of isolates in this study showing subtype *ayw2* HBV genotype D, compared with concensus sequence.[13]

The phylogenetic tree, constructed by comparing fifteen sequences of this study with 58 sequences retrieved from GenBank, confirmed the genotype of our isolates as D. The tree made different clusters of genotype A, B, C, D, E, F, G, and H (Figure 1) and all 15 isolates clustered with D sequences.

The phylogenetic tree obtained by the neighbour-joining method was based on 967 nucleotides of the S gene. Thirteen of the samples (87%) clustered on one branch with genotype D sequences from China, Mongolia, India and Italy. The D strains from Pakistan differed with 0–14 nucleotides from each other, mean 6 nucleotides difference. The 15 HBV isolates demonstrated considerable genotypic heterogeneity and intermixed with strains from India, Middle East and Europe.

The comparison of 15 isolates of large S protein with a consensus sequence showed serotype *ayw2* for all isolates. Twelve of the strains had identical S gene amino acid sequences. All expressed the genotype D specific residues T^{45}, T^{46} and Y^{134} (Figure 2). All strains had specific amino acid substitution F (phenylalanine) at 161 in antigenic determinant of surface gene as well as R (arginine) and P (proline) at 122 and 127 specific for *a* determinant. The amino acid residues P and K (lysine) specifying *y* and *w* were present at positions 127 and 160 and were thus specifying *ayw*.[12] Residue 127, important for the sub-determinants of *w,* was P in all samples, thus specifying *ayw2*.[4] Other determinants for *ayw2* are F^{85}, I^{110}, S^{114}, T^{126}, T^{131}, Y^{134}, T^{140}, S^{143}, F^{158}, G^{159}, K^{160}, F^{161} and A^{168}. Strains EF584652 had a V14A substitution, strain EF584648 had a N40K and strain EF584841 had a L77M substitution.

DISCUSSION

This study provided comprehensive information on the occurrence and composition of HBV genotypes in the studied group. Genotype D was found to be the most prevalent amongst the population studied, infecting 64% of the patients, whereas, genotype A was next most prevalent with 23% incidence and a combination of A and D in 13% of the patients. In an earlier report on comparison of the clinical outcome of HBV-related liver disease in patients infected with different HBV genotypes. Genotype D was the most prevalent (70.9%), in all categories of patients suffering from hepatic disease due to HBV (acute, chronic, carriers and cirrhosis/HCC), whereas, genotype A (20.2%) and AD (9.1%) were next most prevalent, respectively.[14]

Genotype D happens to be the most widely distributed genotype and has been found universally, with its highest prevalence in a belt stretching from Southern Europe and North Africa to India, and in West and South Africa having intravenous drug users as the most affected population within these countries.[15,16] Regarding the prevalence of HBV infection in Pakistan, multiple studies based on the various study design, population selected, diagnostic assays and demographical and epidemiological variations, present a varying rate of infection. According to the study groups, the HBV prevalence rate has been reported as 2-10% among healthy blood donors; 5-9% among health care personnel; 3.6-18.6% among the general population; 3.1% among the pregnant women; 10-20% in patients with provisional diagnosis of hepatitis and 3.1-10.4% among professional blood donors.[17] Pakistan has been reported as among the many regions with high HBV sero-prevalence, having D as most prevalent genotype.[7] Earlier reports published regarding genotype distribution in Pakistan appears to be in contrast to this. Idrees *et al.* reported prevalence of 4 genotypes A, B, C and D with varying degrees of predominance in different provinces of Pakistan. Genotype A (68%) is predominant in Sindh; genotype C in North-West Frontier Province (N.W.F.P.) (68.9%) whereas, genotypes B and C with more or less even dominance in Punjab (39.6% and 25.8% respectively).[18] Later on, other studies on HBV genotyping demonstrated genotype D as the most prevalent in Pakistan,[7] which is confirmed in this study.

Since in Asia, initial research on genotyping was conducted extensively in Japan and China, therefore, B and C genotypes of these countries were considered as the most prevalent genotypes of Asia. Later on, it was found that all the seven HBV genotypes can be found in Asia, with predominance of D in South Asia and Middle East such as in India, Afghanistan and Iran.[19-22] Thus, pattern of genotype prevalence in Pakistan is in accordance with studies from South-East Asia, especially countries sharing borders with Pakistan such as Afghanistan, Iran and India having dominance of genotype D.

The second most prevalent HBV genotype in this study was genotype A (20%) and combination of AD (10%). Studies from India have generally reported a similar pattern of predominance of genotype D coexisting with A and AD in lesser percentage. This raises the same possibility as that of Indians that the Pakistani population originally had genotype D, which has been replaced by genotype A due to human migration from Europe over the time.[23]

Recombination of genotype A and D is also a common event in countries, where different genotypes are prevalent. These co-circulating genotypes provide a mechanism of variation within individuals and in the population in general. Genotype A and D recombinants have been reported from Africa and C/D recombinants from Tibet.[24,25] In the present study, however, no evidence of recombination was found and neither confirmed by phylogenetic analysis. Since genotype A and D are confined to both overlapping and non-overlapping regions of the P/S open reading frames, fewer recombinants were found between these genotypes.[24] The mechanism for this recombination and switching remains an enigma. This could be a possible long-term sequelae leading to reorganization of nucleotides.

In the present study, the genomic group of HBV DNA in sera of diverse geographical origins was assessed by sequencing the viral S gene. The phylogenetic tree obtained by the neighbour-joining method was based on 967 nucleotides of the S gene. Thirteen of the samples (87%) clustered on one branch with genotype D sequences from China, Mongolia, India and Italy. The D strains from Pakistan differed with 0-14 nucleotides from each other, with a mean 6 nucleotides difference. It seems that the HBV isolates of this study demonstrate considerable genotypic heterogeneity as that of the Indian genotype D, and intermix with strains from India, Middle East and Europe.[26]

The HBV subtypes are mainly clustering of HBV DNA sequences into genotype groups corresponded to their respective subtype, that is, *adw2* in genotype B, *adr* in genotype C and *ayw* in genotype D. The “a” determinant is common for all serotypes.[12] Mutations in the "a" determinant may lead to difference in immune-responsiveness, as in vaccine escape mutants and in strain from patients treated with immune globulins.[27] The major serotypes of genotype D are *ayw2, ayw3* and to a lesser extent are *adw2* and *ayw4*. The serotype specified by all our samples with genotype D was *ayw2* compared with the strain pHBV3200 within genotype A.[13] All samples had specific amino acid substitution Y

at codon 134, F at 161 and A at 168 in antigenic determinant of surface gene as well as R and P at 122 and 127 specific for *a* determinant (Figure 2).

In another study on subgenotyping of 13 HBV sequences (unpublished data) D1 was observed as the major subgenotype, whereas a couple of strains as D3. Two of the strains, EF584643 and EF584644, were identical, derived from a phylogenetically related common ancestor. These two D strains clustered on a branch within the D3 samples and were more similar to each other than to the other D3 strains. They differed with 7 nucleotides from each other and derived from a phylogenetically related common ancestor. More studies are needed to investigate the sub-genotypes of the D strains and to find whether they have had fewer introductions into Pakistani strains. Strain EF584642 was the most divergent strain with 4-13 nucleotides difference from the other D strains from Pakistan.

CONCLUSION

Phylogenetic analysis confirmed the dominance of genotype D subtype *ayw2* in Pakistan and advocated a close association with HBV strains that circulate in Iran, India and Japan. More studies are required to find out the mutational frequency in the hepatitis B virus in our environment.

Acknowledgement: This work is dedicated to late Professor Dr. Sarwar Jehan Zuberi, who conceived the idea and supervised the initial part of this work.

REFERENCES

1. Norder H, Courouce AM, Coursaget P, Echevarria JM, Lee SD,Mushahwar IK, Robertson BH, et al. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 2004; **47**: 289-309.
2. Kay A, Zoulim F. Hepatitis B virus genetix variabilty and evolution. *Virus Res* 2007; **3**: 23-9.
3. Chu CJ, Lok AS. Clinical significance of hepatitis B virus genotypes. *Hepatology* 2002; **35**:1274-6.
4. Norder H, Couroucé AM, Magnius LO. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* 1994; **198**:489-503.
5. Zuberi SJ, Lodi TZ. Sub-specificities ob hepatitis Bs antigen. *J Pak Med Assoc* 1978; **28**:1-8.
6. KazmiK, Burney M.I. Subtypes of HbsAg studies in Islamabad. *Pak J Med Res* 1980; **19**:36-8.
7. Alam MM, Zaidi SZ, Shaukat S, *et al.* Common genotypes of hepatitis B virus prevalent in injecting drug abusers (addicts) of North-West Frontier Province of Pakistan. *Virology J* 2007; **4**: 63-77.
8. Naito H, Hayashi S, Abe K. Rapid and specific genotyping system for hepatitis B virus Corresponding to six major genotypes by PCR using type-specific primers. Available from: www.biomed.experts.com.
9. Thompson JD, Higgins DG, Gibson TJ, CLUSTAL W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**:4673-80.
10. Kumar A, Kumar SI, Pandey R, Naik S, Aggarwal R. Hepatitis B virus genotype A is more often associated with severe liver disease in Northern India than is genotype D. *Indian J Gastroenterol* 2005; **24**:19-22.
11. Felsenstein, J. PHYLIP inference package version 3.5. Department of Genetics, University of Washington, Seattle,1993.
12. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, *et al.* Typing hepatitis B virus by homology in nucleotide sequence: comparison of Surface antigen subtypes. *J Gen Virol* 1988; 69: 2575-83.
13. Valenzuela P, Quiroga M, Zalvidar J, Gray P, Rutter WJ. The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. In: Fields BN, Jaenisch R, Fox, edi. Animal virus genetics. New York: *Academic Press* 1980: 57-70.
14. Baig S, Siddiqui A, Ahmed W, Qureshi H, Arif A. The association of complex liver disorders with HBV genotypes prevalent in Pakistan. *Virology* 2007; **4**:128- 32.
15. Borchani-Chabchoub I, Gargouri A, Mokdad-Gargouri R, Genotyping of Tunisian hepatitis B virus isolates based on the sequencing of preS2 and S regions. *Microbes Infect* 2000; **2**: 607-12.
16. Flodgren E, Bengtsson S, Knutsson M, Strebkova EA, Kidd AH, Alexeyev , *et al.* Recent high incidence of fulminant hepatitis in Samara, Russia: molecular analysis of prevailing hepatitis B and D virus strains. *J Clin Microbiol* 2000; **38**:3311-6.
17. Shah NH,. Shabbir G. A review of published literature on hepatitis B and C virus prevalence in Pakistan. *J Coll Physicians Surg Pak* 2002, **12**:368-71.
18. Idrees M, Khan S, Riazuddin S. Common genotypes of hepatitis B virus. *J Coll Physicians Surg Pak* 2004; **14**:344-7.
19. Toan NL, Song LH, Kremsner PG, Duy DN, Binh VQ, Koeberlein B, *et al.* Impact of hepatitis B virus genotypes and genotype mixtures on the course of liver disease in Vietnam. *Hepatology* 2006; **25**:1375-84.
20. Thakur V, Guptan RC, Kazim SN, Malhotra V, Sarin SK. Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in the Indian subcontinent. *J Gastroenterol Hepatol* 2002; **17**:165-70.
21. Amini Bavil-Olyaee S, Sarrami-Forooshani R, Mahboudi F, Sabahi F, Adeli A, Noorinayer B, *et al.* Genotype characterization and phylogenetic analysis of hepatitis B virus isolates from Iranian patients. *J Med Virol* 2005; **75**:227-34.
22. Amini Baril-Olyaee S, Alavian SM, Adeli A, Sarrami-Forooshani R, Sabahi F, *et al.* Hepatitis B virus genotyping, core promoter and precore/core mutations among Afghan patients infected with hepatitis B: a preliminary report. *J Med Virol* 2006; **78**:358-64.
23. Chowdhury A, Santra A, Chakravorty R. Community-based epidemiology of hepatitis B virus infection in West Bengal, India: prevalence of hepatitis B e antigen-negative infection and associated viral variants. *J Gastroenterol Hepatol* 2005; **20**:1711-6.
24. Owiredu WK, Kramvis A, Kew MC. Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B

surface antibody alone: possible association with recombination between genotypes A and D. *J Med Virol* 2001; **64**:441-54.

25. Cui C, Shi J, Hui L. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol* 2002; **83**:2773-7.
26. Banerjee A, Chandra PK, Datta S, Biswas A, Bhattacharya P, Chakraborty S, *et al.* Frequency and significance of hepatitis B virus surface gene variant circulating among antiHBc only individuals in Eastern India. *J Clin Virol* 2007; **40**:312-22.
27. Ruiz-Tachiquín ME, Valdez-Salazar HA, Juárez-Barreto V, Dehesa-Violante M, Torres J, Muñoz-Hernández O. Molecular analysis of hepatitis B virus "a" determinant in asymptomatic and symptomatic Mexican carriers. *Virol J* 2007; **4**:6-11.