Hepatitis B Virus Genotype D is the Only Genotype Circulating in Iranian Chronic Carriers, the Unique Pattern of Genotypic Homogeneity

ABSTRACT

AIM: To characterize the hepatitis B virus surface protein genotypes and sequence variations among HBsAg positive chronic Iranian patients from different ethnic groups.

METHOD: The surface genes from 312 patients were amplified and directly sequenced.

RESULTS: All strains (100%) belonged to genotype D and subtype ayw2. The average nucleotide mutation frequency was 0.91 (dN/dS < 1.0), indicated negative selection. There was no significant correlation between HBV DNA and ALT levels and the occurrence of amino acid substitutions. However, in terms of HBeAg/Anti-HBe status, the association between both groups for silent nucleotide mutation was strong, indicating selection bias on missense mutations. A higher number of amino acid mutations was found in anti-HBe positive versus HBeAg positive patients.
CONCLUSION: The uniqueness pattern of HBV genetics homogeneity together with the low mutational frequency indicated that HBV has introduced to Iran recently and isolation of people in the absence of intermixing with other genotypes led to a homologous pattern.

© 2014 ACT. All rights reserved.

Key words: HBV genotype D; HBV genotypes evolution; HBV surface mutations


INTRODUCTION

The majority of the morbidity and mortality associated with hepatitis B is manifested in conditions, particularly cirrhosis and hepatocellular carcinoma (HCC) as complications of chronic infection. Globally, 30% of the former and 53% of the latter are attributed to hepatitis B infection, respectively.[1]

The HBV surface protein (HBsAg) changes have been classified as either “mutations” (arose after vaccine/drug therapy) or “variants” (determined by host HLA amino acid arrangement over a long period).[2] According to the former classification, HBV mutations define a phenomenon that has been attributed to be the potential mechanism for the pathogenesis basis of chronicity and the clinical complications of this infection. Thanks to the invention of polymerase chain reaction (PCR) and other facilities for direct sequencing and other molecular approaches, hundreds of reports have been published so far to reveal the relationship between correspondent mutations and the clinical/serological pictures of the chronic patients.

Based on the above classification, HBV genome variability can usefully be classified into at least eight families (genotypes) based on sequence divergence in the entire genome of >8-17[3-6]. Further, variation within a sub-component of the S gene within the major hydrophilic region (MHR) of HBsAg, the “a determinant”, is strongly associated with subtype variation[7].

Hepatitis B virus genotypes show a characteristic geographic distribution with a proposed association with human migration. Among HBV genotypes, D is the most prevalent and the most distributed genotype that has been found worldwide. It is found in Western populations, the Indian subcontinent, the Middle East and North of Africa. Genotype D contains 9 subgroups (D1 to D9), based on an intragroup divergence of less than 4% with a global distribution[8], along with 3 subtypes (ayw2, ayw3 and ayw4). Moreover, genotype D has been reported as the only prevalent genotype infecting IV drug users in the Western World[9].

Recent publication from different parts of Iran indicated that genotype D is the only detected HBV genotype[10-14]. In this study, we aimed to analyze the surface gene and protein ina large group of patients who were of various geographical origins and ethnic groups.

METHODS

Study Designs

The study population comprised of sera collected between 2006-2010. The major inclusion criterion was that a specimen should have been tested for HBsAg more than 6 months, with levels of ALT around the normal range. All patients were negative for antibodies against hepatitis C, hepatitis D and human immunodeficiency virus. To avoid bias on the mutational analysis, the patients needed to be treatment and vaccine naïve.

Study Subjects

312 HBsAg-positive patients who were referred to the Iranian Hepatitis Network (IHN) (2006-2010), were enrolled in a cross-sectional study. To cover the whole ethnic parts of country, we studied different regions based on ethnic and geographical zones (Table 1). These patients were belonged to a cohort group whose details of HBV molecular analysis already had been published in recent years[15-20] on the basis of individual province study in connection with IHN. The capital city, Tehran, was chosen as a versatile multi-cultural and multi-ethnic region without a definite ethnic background. Prior informed consent was obtained from all patients before bleeding. 5 mL aliquots of whole blood samples were withdrawn from each participant. Sera were separated, then transferred to Hepatitis B laboratory at Tehran University of Medical Sciences until tested. In our laboratory, HBV serological markers including HBsAg and HBcAg/anti-HBe were examined by ELISA kits manufactured by Organon Technika, Holland.

Table 1 Origin of 312 Iranian HBsAg positive sera that were used as the source for HBsAg sequencing. The submitted sequences to GenBank are listed under accession numbers.

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Region</th>
<th>Number of samples</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkish</td>
<td>(North-west)</td>
<td>17</td>
<td>HM348519-35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GU938342-61, 63-64</td>
</tr>
<tr>
<td>Khorasan</td>
<td>(North-East)</td>
<td>37</td>
<td>HM348519-35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HCQ00867-68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KC176161-72</td>
</tr>
<tr>
<td>Gilae</td>
<td>(North)</td>
<td>5</td>
<td>KC176137-41</td>
</tr>
<tr>
<td>Kurdish</td>
<td>(West)</td>
<td>46</td>
<td>HM348636-81</td>
</tr>
<tr>
<td>Fars</td>
<td>(South)</td>
<td>19</td>
<td>KC176142-60</td>
</tr>
<tr>
<td>Hormozgan</td>
<td>(South)</td>
<td>17</td>
<td>GU938305-12, 14-22</td>
</tr>
<tr>
<td>Baloch</td>
<td>(South-East)</td>
<td>21</td>
<td>HM348694-714</td>
</tr>
<tr>
<td>Khozestan</td>
<td>(South-West)</td>
<td>12</td>
<td>HM348682-93</td>
</tr>
<tr>
<td>Isfahan</td>
<td>(Center)</td>
<td>19</td>
<td>GU938312-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KC176076-99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KC176100-30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HM358277-99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HM358300-29</td>
</tr>
<tr>
<td>Tehran</td>
<td>(Center, Capital)</td>
<td>119</td>
<td>HM358355-39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KC176131-36</td>
</tr>
</tbody>
</table>

DNA extraction and Polymerase Chain Reaction

HBV DNA was extracted from a 200 μL of sera using Qiagen Mini Blood Kit (Qiagen, Hilden, Germany) with a procedure adapted from the manufacturer’s instructions. DNA was eluted using 100 μL of elution buffer, stored in -20°C. Surface gene including ‘a’ determinant antigenic domain sequences was chosen for amplification in order to determine HBV genotype/subtype and other mutations as described previously[21].

DNA sequencing and sequence analysis

Direct sequencing of surface genes was carried out (Perkin Elmer
ABI-3130XL DNA Sequencer, Fostercity, CA, USA) using 0.5 μL of appropriate primers S6 and S7 for surface gene. The results were analyzed using Chromas (version 2.1.1.) and BioEdit (version 7.0.5.3.) software. The HBsAg genotype/subtype of the sequences was defined by substitutions in the ‘a’ determinant between codons 122 and 160 inclusive.

With regards to adequate number of Iranian taxa deposited in GenBank and NCBI, for sequencing alignment comparison, any amino acid differences between isolates defined as “mutation”. All sequence analysis procedures were performed twice by two independent persons for a more precise analysis. Sequences have been submitted to GenBank, numbered according to the time of submission (Table 1).

### Phylogenetic Analysis

Phylogenetic analysis was performed and two neighbor-joining phylogenetic trees were constructed using the MEGA5 employing a Kimura distance matrix. In the first tree, all 312 isolates were included, whereas in the second tree, only 22 isolates that were mutation-free (either at the levels of nucleotide or amino acid) were chosen. A genotype E sequence was chosen for outgrouping (accession number AB032431).

### RESULTS

312 HBsAg-positive chronic patients were enrolled in this study, which all were native residents between different regions of Iran. All were chronic carriers, HBV DNA positive and treatment-naive. 221 (70.83%) were male and 91 (29.16%) were female with a mean age of 36.3±14.01 (mean±SD). The mean ALT and viral load level were 85±109.2 (mean±SD) and 15,000 copy/mL, respectively (results not shown) 23.39% (n=73) and 70.83% (n=221) were HBeAg and anti-HBe positive, respectively. Also 5.12% (n=16) and 0.64% (n=2) were negative and positive for both markers, respectively.

**Phylogenetic Analysis**

The results of the first large phylogenetic tree (contained all the studied-sequences) revealed that Iranian HBV isolates were of genotype D, supported by 95% bootstrap value (1,000 replicates). A majority of isolates clustered with the Caucasian origin isolates; from India, Germany and Italy (due to the complexity of tree the results are not shown). Finally, all isolates were belonged to subtype ayw2.

**Inter-group substitutions in comparison with reference genotype D (Okamoto, AB033559)**

The results of the second phylogenetic tree that was designed using only 22 sequences with no nucleotide and amino acid mutations (Figure 1), all strains clustered in a separate branch from all other genotypes D. The most comparable genotype D isolate to Iranian sequences was AB033559 that was originated from Papua New Guinea. Overall, comparing with this reference sequence, at the nucleotide level, six changes were occurred (G208C, A339T, C360A, A420C, C438T and T613C) of

![Figure 1 Phylogenetic tree of 22 Iranian surface genes sequences from nucleotide and amino acid mutation-free samples. Note: S gene tree rooted with genotype E (accession number AB032431). The scale denotes percent diversity. Coding numbers indicate samples that have been analysed in the figure. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed.](image-url)
which, G208C was a non-synonymous substitution that at the amino acid levels altered alanine to proline (A70P) (results not shown). We believe that this substitution was assigned as “variant” (see material and methods). According to the above mentioned description, 1,058 and 505 changes at the nucleotide and amino acid levels were true mutations, respectively (results not shown).

### Intra-group nucleotide and amino acid substitutions in comparison with Iranian sequences

In comparison with Iranian sequences obtained from the database as well as from our unpublished data, in addition to the genotypic characterization described above, the sequences of the strains showed a few variability over the regions sequenced. In total, 205 (65.70%) of 312 patients contained at least one amino acid mutation (Table 2). 85 (27.24%) of patients contained only silent nucleotide mutations with no subsequent amino acid alteration (Table 2). 22 (7.05%) of patients did not contain any nucleotide and amino acid substitutions at all. In all, 1,058 “nucleotide mutations” occurred, of them 505 (47.73%) were missense (amino acid altering) and 553 (52.26%) were silent (no amino acid changing). The average nucleotide mutation frequency was 3.6 events per sample. There was a probability of 0.5% for substitution per nucleotide position. Further, there were 32 non-sense nucleotide substitutions (leading to a stop codon at the amino acid level). The average amino acid mutation frequency was 1.68 mutations per sample. There was a probability of 0.74% for substitution per amino acid position. In addition, it was possible to identify the level of S proteins evolution between isolates by measuring the mutation frequency (dN/dS; Missense/Silent ratio) of individual sequences. The average mutation frequency of all sequences was 0.91 according to the number of mutations per site. The details of surface protein amino acid mutations distribution had been published previously elsewhere[24].

### Surface protein variations and clinical status

The comparison between the distribution of amino acid substitutions in the surface proteins with age, gender and ALT levels of patients showed no significant association (results not shown). Similarly, in total, there was no significant correlation between HBV DNA levels (<10^4 versus >10^4 copies/mL) and the occurrence of amino acid substitutions (results not shown). Nonetheless, in terms of HBeAg/Anti-HBe status, when the total number of nucleotide variations compared between both groups, the correlation was not significant (Table 2). 22 (7.05%) of patients did not contain any nucleotide and amino acid substitutions at all. In all, 1,058 “nucleotide mutations” occurred, of them 505 (47.73%) were missense (amino acid altering) and 553 (52.26%) were silent (no amino acid changing). The average nucleotide mutation frequency was 3.6 events per sample. There was a probability of 0.5% for substitution per nucleotide position. Further, there were 32 non-sense nucleotide substitutions (leading to a stop codon at the amino acid level). The average amino acid mutation frequency was 1.68 mutations per sample. There was a probability of 0.74% for substitution per amino acid position. In addition, it was possible to identify the level of S proteins evolution between isolates by measuring the mutation frequency (dN/dS; Missense/Silent ratio) of individual sequences. The average mutation frequency of all sequences was 0.91 according to the number of mutations per site. The details of surface protein amino acid mutations distribution had been published previously elsewhere[24].

### Table 2 Pattern of nucleotide/amino acid distributions within the surface protein in correlation with HBeAg/Anti-HBe status of the patients.

<table>
<thead>
<tr>
<th>Substitution levels</th>
<th>HBsAg Positive (n=73)</th>
<th>Anti-HBe Positive (n=221)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide mutation</td>
<td>65</td>
<td>200</td>
<td>0.82</td>
</tr>
<tr>
<td>Nucleotide Silent Mutations (n=85)</td>
<td>31</td>
<td>49</td>
<td>0.001</td>
</tr>
<tr>
<td>Amino Acid Mutation</td>
<td>36</td>
<td>162</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>“a” determinant Mutations (n=34)</td>
<td>7</td>
<td>27</td>
<td>0.67</td>
</tr>
<tr>
<td>No nucleotide and amino acid Mutations (n=22)</td>
<td>7</td>
<td>12</td>
<td>0.27</td>
</tr>
</tbody>
</table>

### Discussion

It is important that the prevalence and types of variants of the S gene found in populations to be recorded, because this will affect policy decisions relating to vaccine and diagnostic reagents design. Therefore, in the present study, genetic variations in the whole HBsAg were analyzed comprehensively in 312 chronic patients from 9 different parts of Iran. The patient samples included in this study represent selected material from all major parts of the country with the largest sample size (in terms of HBV genotyping so far) including the main ethnic groups between Iranian populations.

Our comparison of 312 new complete surface genes and proteins with those in databases verified that genotype D and subtype ayw2 accounted for 100% of isolates, respectively. Published data indicated that in Iran, there has been an obvious predominance of these virus genetic patterns[10-13]. Other unpublished data (from personal communications) indicates that the genotype D is the only genotype of hepatitis B circulating in this country. Therefore, according to epidemiological studies so far, Iran is the only region in the world which harbored the unique pattern of HBV genotypic pattern.

According to archaeological and anthropological findings, the ancestors of Caucasians (Arians) firstly colonized to the North of the Caspian Sea. Because of difficulties in agriculture and climate change, they migrated in three directions thousands years ago: one group moved west towards Europe, another group moved south to Iran (established the ancient Persian Empire) and the last group migrated to India. It might be that those people, who acquired the virus with the genotype D before their migration [perhaps related to immune pressure based largely on human leukocyte antigens (HLA) types], then transmitted the virus generation by generation after their migration. This hypothesis is underscored by the findings that the dominant Caucasians HBV genotype in India, Iran and most part of the Europe is D (Figure 1). After colonization of infected people, certain genotypes importation (especially B and C) to this area arose from miscegenation of people. Further evolution occurred, giving some of these sequences a distinctive motif, and some genetic recombination between genotypes also occurred, which led to the heterogenous pattern like India, Pakistan[14]. In some areas, isolation of people in the absence of intermixing with other genotypes let to a homogonous pattern (like Iran and Turkey)[14]. Thus, the unique pattern of genotypic homogeneity in Iran reflects a founder event, in which, a new population is founded by a small number of incoming individuals, giving some of these sequences a distinctive motif, and some genetic diversity, leading to homogeneity. The latter finding indicates a unique homogeneity belonged to the Iranian population.

The ratio between silent and missense nucleotide mutations (dN/dS) in our patients was 0.91 indicating that the proportion of deduced amino acid changes in these chronically infected patients followed a negative selection pattern. A conclusive assumption needed by comparing the whole HBV genome sequences, however, according to the average amino acid substitution and nucleotide frequency, we hypothesize that the evolution of HBV surface protein between Iranian patients is somehow low. Based on current investigations on genotype D of HBV, new findings show that HBV genotype D is able to cause severe diseases and higher rate of drug resistance in comparison to other studied genotypes. Thakur et al[20] reported...
that genotype D of HBV correlates with more severe liver disease than HBV genotype A in India, especially in young HBV infected patients may be lead to HCC. Other report comparing the clinical outcome of chronic HBV infection between different genotypes indicating the worst complications caused by genotype D[25-27]. We were not being able to compare our group of patients with genotype D with the other groups of HBV genotypes. However, according to the epidemiologic studies, the prevalence of cirrhosis and HCC, the major complications of chronic HBV infection in Iran (including these regions-studied) are relatively low[20-29] and HCC is not included in the list of top ten cancers in Iranian population (regardless of role of HBV as an etiology)[30-32]. In chronic carriers, the transition from a relative immune tolerance state (HBsAg positivity) to the activation of the immune system with generation of anti-HBs results in a strong selection on the viral genome, causing changes in immune targets (HBsAg). This seroconversion reflects the evolution of the accumulated mutations as time goes by according to molecular clock. Although the impact of such mutations in the pathogenesis of cirrhosis and HCC is not clear, the rough conclusion is that due to this negative selection and the uniqueness pattern of HBV genotype in this ethnic group, these complications are lower than other published studies so far[33]. A definite conclusion needs mutational analysis of sequential samples from different stages of chronically-infected individuals in cohort groups of different genotypes.

In conclusion, the unique pattern of HBV genetics homogeniety together with the low mutational frequency indicated that HBV has introduced to Iran recently and isolation of people in the absence of intermixing with other genotypes let to a homologous pattern. This uniqueness deserves more epidemiological and biomolecular studies with regard to other genotypes to explore the biological impacts of different strains on the complications of HBV infection.

ACKNOWLEDGMENTS

This study was supported by Tehran University of Medical Sciences (Grant No: 860227-5816). We are deeply indebted to Dr Helen Norder and Professor Lars Magnus for technical assistance of phylogenetic tree designing.

CONFLICT OF INTERESTS

There are no conflicts of interest with regard to the present study.

REFERENCES


Norouzi M et al. Exclusive distribution of HBV genotype D in Iran

28 Alavian SM. Hepatitis B virus infection in Iran; Changing the epidemiology. *Iranian Journal of Clinical Infectious Diseases* 2010; 5(1): 51-61

Peer reviewer: Mohammad Khalid Pravez, PhD, Department of Pharmacognosy, King Saud University College of Pharmacy, Riyadh 11541, Saudi Arabia.