COMPLETE GENOME SEQUENCES AND PHYLOGENETIC ANALYSIS OF HEPATITIS B VIRUS ISOLATES FROM SERBIA

IVANA LAZAREVIĆ1, MAJA ĆUPIĆ1, D. DELIĆ2, NEDA STOJKOVIĆ-SVIRTLIH2, JAMINA SIMONOVIĆ2, ANA BANKO1 and TANJA JOVANOVIĆ1

1 Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, 11000 Belgrade, Serbia
2 Infectious and Tropical Diseases Clinics, Faculty of Medicine, University of Belgrade, 11000 Belgrade, Serbia

Abstract - Although the genotype and subgenotype distribution of HBV isolates from Serbia has previously been reported, data about whole genome sequences from this area are scarce. This study included plasma samples from 5 chronically infected patients. Full genome amplification of the HBV isolates was performed by nested-PCR using 7 primer pairs, and the whole genome nucleotide sequences were obtained by direct sequencing. Two complete genome sequences belonged to D2 subgenotype (ayw3 HBsAg subtype), one to D1 (ayw2) and two to A2 (adw2). All 5 Serbian isolates clustered with sequences from the expected geographic regions and had nucleotide and coded protein length in accordance to their assigned genotypes, except for one HBeAg-negative isolate displaying G1896A mutation leading to a premature stop codon in the Pre-C region. The first complete genome sequences of HBV D1, D2 and A2 subgenotypes from Serbian patients showed characteristics similar to the nucleotide sequences of HBV isolates from other European and Middle East countries.

Key words: Hepatitis B virus (HBV), genome, genotype, phylogenetic analysis

INTRODUCTION

Hepatitis B virus (HBV) is an important etiologic agent of acute or chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. It is estimated that there are currently more than 350 million chronically infected people worldwide.

Since HBV replication involves an error-prone reverse transcription step, the HBV genome evolves quickly over time, with an estimated nucleotide substitution rate between 1.5 and 7.9 x 10⁻⁵ substitution per site per year. This unique replication strategy accounts for the majority of the point mutations, deletions and insertions observed in the HBV genome. The lengthy evolution of HBV has led to the current existence of various genotypes, subgenotypes, mutants, recombinants, and even quasispecies of HBV (Lee and Locarnini, 2004; Osiowy et al., 2006).

According to the homogeneity of virus sequences, 10 HBV genotypes (A to J) have been defined by divergence in the entire HBV genomic sequences of >8%. Except for the newly identified genotypes I and J, the geographic and ethnic distributions of HBV genotypes and subtypes are well characterized. Genotype A is predominant in northwestern Europe and North America; genotypes B and C are confined to the populations with origins in Eastern Asia and the Far East (Okamoto et al., 1988); genotype D is found worldwide but prevails in the Mediterranean area and the Near and Middle East, extending into India (Norder et al., 2004., Amini-
Bavil-Olyaee et al., 2005); genotype E is indigenous to western sub-Saharan areas; genotype F is typical for American natives and Polynesia (Norder et al., 2004); genotype G has been reported in the USA, Germany and France (Stuyver et al., 2000); the eighth genotype, H (Arauz-Ruiz et al., 2002), is found in Central America and recently, genotype I (Phung et al., 2010), a novel inter-genotypic recombination was isolated in Vietnam and Laos; the newest HBV genotype, J (Tatematsu et al., 2009), was identified in Japan.

Further extensive phylogenetic analyses of the HBV genotypes have resulted in recognition of subgenotypes of genotypes A, B, C, D and F, based on more than 4% intragenotypic divergence (Norder et al., 2004). To date, the presence of 7 subgenotypes has been recognized for HBV genotypes B and D (Sakamoto et al., 2007; Meldal et al., 2009), 5 for genotypes A and C (Sakamoto et al., 2006; Olinger et al., 2006), while 4 subgenotypes have been reported for genotype F. The data about the number of new subgenotypes vary due to different taxonomic criteria and more epidemiological, viral and clinical features are still needed for some of the newly described subgenotypes to justify their classification.

Serbia has an intermediate pattern of HBV infection, with a prevalence of HBsAg positivity of approximately 2-7% of the population, and mixed patterns of infant, early childhood and adult transmission. Two HBV genotypes, A and D, and 4 subgenotypes A2, D1, D2, and D3 have been encountered in Serbia, with genotype D and subgenotype D3 being prevalent (Lazarević et al., 2010). So far, only three complete HBV genome sequences of subgenotype D3 from Serbian isolates have been published (Stanoević et al., 2011), while complete genome sequences of other subgenotypes circulating in Serbia have not yet been characterized.

The aim of the present study was to characterize the complete nucleotide sequences of HBV isolates of genotypes A and D from chronically infected patients in Serbia.

**MATERIALS AND METHODS**

The study included 5 plasma samples from patients with chronic hepatitis B, treated at the Clinics of Infectious and Tropical Diseases in Belgrade. The diagnosis was based on HBsAg positivity for >6 months and histopathological verification by liver biopsy. All patients had high viral loads (9.63-11.43 log copies/mL) and all but one were HBeAg-positive. None received any antiviral therapy prior to investigation. The HBeAg-positive and therapy-naïve patients were selected in order to avoid therapy- and immune-derived nucleotide substitutions within the genome. Based on the partial genome sequencing, 3 isolates belonged to genotype D, and 2 to genotype A. The demographic and viral characteristics of the patients/isolates are summarized in Table 1. The study was approved by the Ethics Committee of School of Medicine, University of Belgrade, and informed consent to participate in this study was obtained from the patients.

HBV DNA from the plasma was extracted using QIAamp Blood Mini Kit (Qiagen GmBH, Hilden, Germany) according to the manufacturer’s instructions. DNA amplification was carried out by nested-PCR, where primers F1 and R7 were used in the first round of PCR for amplification of the whole genome, and 7 primer pairs (Table 2.) were used in the second round of PCR for amplification of 7 partially overlapping segments. The primers, previously described by Zhu et al. (2007), were adjusted for the best amplification of genotypes A and D, since these were the expected genotypes in the Serbian population. The amplification protocol included: initial denaturation for 5 min at 95°C followed by 40 cycles of amplification (94°C for 40 s, 60°C for 30 s and 72°C for 4 min with a 30 s increase each five cycles for the first round PCR) and a 7 min final extension at 72°C.

The PCR products were purified using the QIAGEN MinElute Purification Kit (Qiagen GmBH), according to the manufacturer’s protocol. For cycle sequencing reactions, second round PCR primers and a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) were used. Se-
Table 1. Demographic and virological characteristics of HBV infected patients included in the study

<table>
<thead>
<tr>
<th>Patient/isolate No.</th>
<th>Age/Sex</th>
<th>HBsAg</th>
<th>HBeAg</th>
<th>Anti-HBe</th>
<th>Viral load (log copies/ml)</th>
<th>Liver biopsy</th>
<th>Antiviral therapy (prior to study)</th>
<th>Subgenotype (based on partial sequencing)</th>
<th>Subgenotype (based on complete genome sequencing)</th>
<th>HBsAg subtype</th>
<th>Genome length</th>
</tr>
</thead>
<tbody>
<tr>
<td>srb176</td>
<td>29/M</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>10.62</td>
<td>Low grade fibrosis</td>
<td>No</td>
<td>A2</td>
<td>A2</td>
<td>adw2</td>
<td>3221bp</td>
</tr>
<tr>
<td>srb200</td>
<td>36/M</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>10.73</td>
<td>Cirrhosis</td>
<td>No</td>
<td>D2</td>
<td>D2</td>
<td>ayw3</td>
<td>3182bp</td>
</tr>
<tr>
<td>srb205</td>
<td>25/M</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>11.43</td>
<td>Low grade fibrosis</td>
<td>No</td>
<td>D3</td>
<td>D1</td>
<td>ayw2</td>
<td>3182bp</td>
</tr>
<tr>
<td>srb207</td>
<td>38/M</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>9.97</td>
<td>Low grade fibrosis</td>
<td>No</td>
<td>D2</td>
<td>D2</td>
<td>ayw3</td>
<td>3182bp</td>
</tr>
<tr>
<td>srb210</td>
<td>34/M</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>9.63</td>
<td>Low grade fibrosis</td>
<td>No</td>
<td>A2</td>
<td>A2</td>
<td>adw2</td>
<td>3221bp</td>
</tr>
</tbody>
</table>

Table 2. Primers used for amplification and sequencing of HBV genome

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Position in the genome</th>
<th>Primer sequence 5'-3'</th>
<th>Direction</th>
<th>Fragment length</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1821–1841</td>
<td>TTT TTC ACC TCT GCC TAA TCA</td>
<td>sense</td>
<td>505 bp</td>
</tr>
<tr>
<td>R1</td>
<td>2300-2325</td>
<td>CCA CCA AAT GCC CCT TTA TCA AC</td>
<td>antisense</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>2031–2050</td>
<td>CTC CTG AGC ATT GTT CAT CT</td>
<td>sense</td>
<td>820 bp</td>
</tr>
<tr>
<td>R2</td>
<td>2831-2850</td>
<td>GGA ACA AGA TCT ACA GCA TG</td>
<td>antisense</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>2469–2490</td>
<td>AAG GTG GGG AAC TTT ACT GGG C</td>
<td>sense</td>
<td>748 bp (genotype A)</td>
</tr>
<tr>
<td>R3</td>
<td>3197-3217</td>
<td>CAT CCT CAG GCC ATG CAG TGG</td>
<td>antisense</td>
<td>712 bp (genotype D)</td>
</tr>
<tr>
<td>F4</td>
<td>3036–3056</td>
<td>TTG GGG TGG AGC CCT CAG GCT</td>
<td>sense</td>
<td>585 bp</td>
</tr>
<tr>
<td>R4</td>
<td>412–433</td>
<td>CCT GCT GCT ATG CCT CAT CTT C</td>
<td>antisense</td>
<td></td>
</tr>
<tr>
<td>F5</td>
<td>300–318</td>
<td>TTG GCC AAA ATT GCC AGT C</td>
<td>sense</td>
<td>698 bp</td>
</tr>
<tr>
<td>R5</td>
<td>979-997</td>
<td>TGG AAA GTA TGT CAA GGA A</td>
<td>antisense</td>
<td></td>
</tr>
<tr>
<td>F6</td>
<td>715–734</td>
<td>CCC CAC TGT TTG GCT TCT AG</td>
<td>sense</td>
<td>745 bp</td>
</tr>
<tr>
<td>R6</td>
<td>1441-1460</td>
<td>CTG AAT CCC GGG GAC GAC CC</td>
<td>antisense</td>
<td></td>
</tr>
<tr>
<td>F7</td>
<td>1186–1205</td>
<td>GCT GAC GCA ACC CCC ACT GG</td>
<td>sense</td>
<td>640 bp</td>
</tr>
<tr>
<td>R7</td>
<td>1806-1825</td>
<td>CCA GCA CCA TGC AAC TTT TT</td>
<td>antisense</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. Phylogenetic analysis of 5 complete HBV genome sequences from Serbia (srb176, srb200, srb205, srb207, srb210) compared with 82 comparative complete genome sequences represented in the phylogenetic tree with subgenotype, Genbank accession numbers and country of origin. Bootstrap confidence values of ≥70% are given.
sequencing was carried out in an automatic sequencer (ABI PRISM 310 Genetic Analyzer; Applied Biosystems). Sense and antisense strands were sequenced, compared and assembled. The 7 obtained sequences for each isolate were assembled into whole genome sequences in MEGA software version 3.1 (Kumar et al., 2004).

The nucleotide sequences were aligned pairwise using ClustalW, and a phylogenetic tree was constructed by the neighbor-joining method and Kimura two-parameter algorithm. Statistical significance of phylogeny was estimated by bootstrap analysis with 1,000 pseudoreplicate datasets. For comparison, reference sequences for 9 known HBV genotypes from the DDBJ/EMBL/GenBank database were used as well as more isolates of genotypes A and D from different parts of the world (total of 106: 4 for A1, 13 for A2, 2 for A3, 3 for B, 5 for C, 24 for D1, 19 for D2, 11 for D3, 3 for D4, 2 for D5, 3 for D6, D7, E, F, G, H and 2 for I).

The viral genotype and subgenotype were identified by phylogenetic analysis. The HBsAg subtype was deduced from the sequence by identifying amino acids at positions 122, 160, 127, 159, and 140. Both of the genotype A isolates displayed a sequence characteristic for the adw2 HBsAg subtype. As expected from known subgenotype/HBsAg subtype associations, the subgenotype D1 isolate belonged to ayw2, while both D2 isolates belonged to ayw3 (Table 1.).

Both genotype A isolates showed the expected length of 3221 bp with a 6-nucleotide (nt) insert at the carboxyl terminus of the core gene (nt 2356-2361). All genotype D isolates (srb200, srb205 and srb207) showed the characteristic 33-nt deletion in the Pre-S1 region (nt 2859-2891) which corresponded to the spacer region of HBV polymerase gene. Accordingly, the deduced 7 proteins M, L, S of the Pre-S/S gene, P, X and Pre-C/C and C were 400, 281, 226, 845, 154, 214 and 185 aa long for genotype A isolates, and 389, 281, 226, 832, 154, 212 and 183 aa for the genotype D isolates.

All isolates were examined for the presence of clinically relevant mutations in the HBsAg, BCP (basal core promoter) region of the X gene, Pre-C and RT (reverse transcriptase) of the P gene in BioEdit (Hall, 1999) and MEGA software.

The complete genome sequences of 5 isolates were analyzed and compared with 106 isolates of genotypes A and D from different parts of the world and with reference sequences for 9 known HBV genotypes from the DDBJ/EMBL/GenBank database. Two of the analyzed isolates belonged to genotype A, and 3 to genotype D. Both genotype A isolates clustered with the representative sequence for the A2 subgenotype in the phylogenetic tree, whereas genotype D isolates belonged to two different subgenotypes: one to D1 and two to D2. The results are summarized in a phylogenetic tree, where only 82 comparative sequences are shown due to limited space (Fig. 1). All 5 Serbian isolates clustered with sequences from expected geographic regions.

HBsAg subtypes were deduced from the sequence by identifying amino acids at positions 122, 160, 127, 159, and 140. Both of the genotype A isolates displayed a sequence characteristic for the adw2 HBsAg subtype. As expected from known subgenotype/HBsAg subtype associations, the subgenotype D1 isolate belonged to ayw2, while both D2 isolates belonged to ayw3 (Table 1.).

Both genotype A isolates showed the expected length of 3221 bp with a 6-nucleotide (nt) insert at the carboxyl terminus of the core gene (nt 2356-2361). All genotype D isolates (srb200, srb205 and srb207) showed the characteristic 33-nt deletion in the Pre-S1 region (nt 2859-2891) which corresponded to the spacer region of HBV polymerase gene. Accordingly, the deduced 7 proteins M, L, S of the Pre-S/S gene, P, X and Pre-C/C and C were 400, 281, 226, 845, 154, 214 and 185 aa long for genotype A isolates, and 389, 281, 226, 832, 154, 212 and 183 aa for the genotype D isolates.

All isolates were examined for the presence of clinically relevant mutations in the HBsAg region, RT region of P gene, BCP region of X gene and Pre-C region. A specific search was conducted for mutations associated with resistance to nucleot(z)ide analogues at positions 80, 166, 169, 173, 180, 181, 184, 200, 204, 207, 215, 236 and 250 of the RT region, but despite detected minor changes in RT in all isolates (particularly in genotype D), none were found on the...
The HBsAg region was examined for the presence of known mutations related to vaccine escape, immunotherapy resistance and diagnostic failure. Changes T118A and P127T were found in 2 isolates but were not regarded as mutations since they represent subgenotype/HBsAg subtype-dependent polymorphisms (Tallo et al., 2008). Three more changes in this region (I110M, A159V, L193S) were registered in 4 isolates, all outside “ä” determinant. In the BCP region of the X gene, analysis included changes associated with the decreased production of HBeAg A1762T, G1764A and related compensatory mutations at positions 1753, 1766 and 1768. Only one genotype D isolate (srb200) displayed compensatory mutation T1753A in this region. Mutation G1896A of Pre-C region leading to stop codon and complete abolishment of HBeAg production was detected in another genotype D isolate (srb205).

DISCUSSION

Serbia’s geographic position on the main route between the Middle East and Central Europe and in close proximity to the Mediterranean basin has influenced HBV genotype and subgenotype distribution. As previously described, two genotypes were encountered (A and D), with D being dominant and with subgenotype distribution with D3 predominance closest to western Mediterranean countries (Lazarević et al., 2010; Zehender et al., 2008). Recently, the full molecular characterization of the D3 subgenotype was reported by Stanojević et al. (2011), while aim of the present study was to obtain the sequence and description of the full genome sequences of subgenotypes D1, D2 and A2 from Serbian isolates.

All 5 Serbian isolates clustered with sequences from the expected geographic regions, those that have significant demographic and historical ties with Serbia, such as central and eastern European countries and Turkey. This is observed for D1 and D2 clusters, where the Serbian isolates belong to different clusters from isolates originating in Asia or the Far East. In the case of A genotype isolates, both of them were clustered with A2 subgenotype isolates, known to be a “European” subgenotype.

Results revealed by phylogenetic analysis of the full genome sequences of 5 isolates were generally in accordance with data about genotypes and subgenotypes previously obtained by partial S gene-sequencing of the same isolates. Only in the case of one isolate, previously identified as D3, and in this study as D1, was subgenotyping by partial sequencing proved to be inaccurate. This method is reliable for the detection of genotypes since the S gene sequence coding for small S protein has many genotype-specific spots, but differentiation between subgenotypes based on sequence clustering in the phylogenetic tree cannot be perfectly accurate when the examined sequence is not long enough. Therefore, subgenotype distribution based on full genome sequences can perhaps show the higher frequency of D1, particularly because this is the prevalent subgenotype in Asia Minor (Norder et al., 2004, Bozdayi et al., 2005), a region with important historical connections to Serbia.

The clinically relevant mutations were not very frequent in the studied sequences since the isolates chosen were from therapy-naïve patients and mostly HBeAg-positive in order to avoid nucleotide changes derived from these factors. The absence of any RT mutation known to be associated with resistance to antiviral therapy was not unexpected since in most cases these mutations need drug-induced selective pressure to develop.

HBeAg-negative chronic hepatitis B evolves in the natural history of chronic HBV infection and is linked with the selection of nonproducing HBeAg but replication-competent HBV mutants. Although it was first believed to prevail in the Mediterranean area, this form of chronic hepatitis, according to recent data, appears to be more common and is gradually becoming the predominant form in all countries (Hadziyannis and Papatheodoridis, 2006). Mutations responsible for HBeAg-negative hepatitis belong to the BCP region of the X gene and Pre-C region, and are very common among the isolates of chronically infected patients in Serbia.
(Lazarević 2008). Only one HBeAg-positive genotype D isolate displayed compensatory mutation T1753A in the BCP region. This mutation, unlike A1762T and G1764A mutations, is not responsible for decreased expression of HBeAg, but usually accompanies it and is associated with an increased replication fitness of the strain. Therefore, the presence of this mutation alone can perhaps be the first phase in transformation of the strain to HBeAg-nonproducing. The other genotype D isolate, originating from an HBeAg-negative patient, showed the very common Pre-C mutation G1896A, which transforms codon 28 into a stop codon and leads to the abolishment of HBeAg production. This mutation is particularly common in genotype D isolates due to the specific nucleotide sequence that allows the stable structure of epsilon pgRNA and effective replication, even in the presence of this mutation (Key and Zoulim, 2007).

Mutations within the S gene, particularly within the region called “a” determinant (aa124–147), caused by selection or natural variation, can affect the antigenicity of HBsAg and lead to false-negative results by assays for HBsAg; or it can have clinical implications, such as evasion of anti-HBV immunoglobulin (HB Ig) therapy or vaccine-induced immunity. Mutations found in the examined isolates were all outside of this region and were not associated with any clinical and diagnostic problems.

The results of this study are in accordance with country’s geographic position, since the first complete genome sequences of D1, D2 and A2 HBV isolates show characteristics similar to HBV isolates from other European and Middle East countries. The results revealed may enable further characterization of HBV strains circulating in the Balkans.

Acknowledgments - This study was supported by Ministry of Science and Technological Development, Republic of Serbia, research Grant No.175073. For technical assistance in performing the study, the authors thank laboratory technicians Gabriela Pavlović and Marija Janković from the Virology Department, Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade.

REFERENCES


