Hepatitis B Virus Genotype D Strains From Estonia Share Sequence Similarity With Strains From Siberia and May Specify ayw4

Tatjana Tallo,1,2 Helene Norder,2 Valentina Tefanova,1 Tõnu Krispin,5 Ludmilla Priimägi,1 Sergey Mukomolov,3 Mikhail Mikhailov,4 and Lars O. Magnius2*

1Department of Virology, National Institute for Health Development, Tallinn, Estonia
2Department of Virology, Swedish Institute for Infectious Disease Control, Stockholm, Sweden
3St. Petersburg Pasteur Institute, St. Petersburg, Russia
4N.F. Gamaleya Institute of Epidemiology and Microbiology, Russian Academy of Medical Science, Moscow, Russia
5Department of Microbiology, University of Tartu, Tartu, Estonia

The genotypes and subtypes of 205 HBV isolates collected during 1989–2002 in Estonia and 14 other regions of the former USSR were determined by sequencing and phylogenetic analysis of the S gene. The in Europe prevailing genotypes, A and D, were also circulating in the whole territory of the former USSR including Estonia and accounted for 18.5 and 81% of the strains, respectively. All genotype A strains specified adw2, and a single genotype C strain specified adrq+. Most genotype D strains specified ayw3 and ayw2, although, three strains from Estonia and Siberia specified ayw4. Due to unique substitutions, Ser122 and Ala127, four strains could not be classified according to the subtype. One strain specifying ayw3 encoded Leu143 and Ala145 and was possibly an immune “escape” mutant. At phylogenetic analysis 93% of the Estonian genotype D strains belonged to a cluster specifying mainly ayw3 and were more similar to isolates from Siberia and the Far-East of Russia than to isolates originating from Central Russia which belonged to another cluster of strains specifying mainly ayw2. This pattern might be explained by part of the Estonian population, has roots east of European Russia, based on linguistic evidence. Eight dominant HBV strains represented by identical S gene sequences were identified, one within genotype A and seven within genotype D, three of which included isolates from Estonia and Siberia. Some of these strains were collected over a period of at least 13 years indicating there are genetically stable variants of HBV that remain conserved over decades. J. Med. Virol. 74:221–227, 2004. © 2004 Wiley-Liss, Inc.

KEY WORDS: genotype; subtype; phylogenetic analysis; former USSR

INTRODUCTION

Hepatitis B virus (HBV) is a partially double-stranded DNA virus of the Hepadnaviridae family with a genome of approximately 3.2 kb. Despite similarity to retroviruses, i.e., the replication through an RNA intermediate involving reverse transcription, HBV is a highly conserved virus due to the partial overlapping of four open reading frames and the lack of non-coding regions [Ganem and Varmus, 1987].

The serological heterogeneity of the HBV surface antigen (HBsAg) has been established by the immunodiffusion. The identification of a common α-determinant, two mutually exclusive determinant pairs, d/y [Le Bouvier, 1971] and w/r [Bancroft et al., 1972], the q-determinant [Magnius et al., 1975] and the subdeterminants of w [Couroucé et al., 1976] enables the classification of HBV isolates into nine subtypes: ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adrq+, and adrq−. The molecular basis for these serological variations has been identified. Thus, variations of the S-gene product at residues 122 and 160 identify the d/y and w/r specifi-
The genetic classification of HBV strains based on an intragroup divergence of less than 4% and intergroup divergence of 8% or more of the complete genomes once defined four genomic groups, A through D [Okamoto et al., 1988]. The divergence between the genomic groups was 4.1% when comparing S-gene sequences and enabled the identification of two more genomic groups now referred to as genotypes E and F, also confirmed by sequencing of their complete genomes [Norder et al., 1992a]. HBV is now classified into eight genotypes designated A to H [Naumann et al., 1993; Norder et al., 1994; Stuyver et al., 2000; Arauz-Ruiz et al., 2002]. Based on the variability of the S-gene a worldwide molecular epidemiology of HBV has been established [Norder et al., 1993]. Thus, genotype A is common in North-Western Europe, North America, and Africa. Genotypes B and C are confined to Asia. Genotype D is the most widely distributed genotype and occurs all over the world with highest prevalence in the Mediterranean, the Near East, and India. E is the predominant genotype in West Africa. G is a genotype described from chronic HBV carriers in North America and France [Stuyver et al., 2000]. Genotype F occurring in Central and South America and genotype H found in Central America, Mexico, and California are considered the original HBV genotypes of Amerindians [Arauz-Ruiz et al., 2002].

There is limited information regarding the genetic variability of HBV in Estonia previously a part of the USSR, as well as the former USSR itself. The aim of this study was, therefore, to determine the distribution of HBV strains in Estonia and to assess their genetic relatedness to strains from other territories of the former USSR.

**MATERIALS AND METHODS**

**Sera**

One hundred thirty four HBsAg positive sera were sampled in Estonia and 396 in 14 other regions of the former USSR, including Latvia, Belarus, Moldavia, Russia (Moscow region, Volga region, Tyumen region, Kemerovo region, Krasnoyarsk region, Chita oblast, Yakutia oblast, Khabarovsk region) and three republics of Middle Asia (Kazakhstan, Uzbekistan, Turkmenistan) (Table I). In all 530 sera were available, 174 were from patients with acute hepatitis B, 26 from persons with chronic hepatitis B, and 330 from blood donors.

**Methods of HBsAg Detection**

Samples collected in Estonia during 1996–2002 were tested by enzyme immunoassay (EIA) with commercial kits for HBsAg detection from Organon Technika, Sanofi Diagnostics Pasteur, Murex Diagnostics, Ortho Diagnostic Systems, and DiaSorin. Sera collected in 1989, 1991, and 1997 in the other territories of the former USSR were investigated by EIA and passive agglutination assays with commercial kits (Diagnostic Systems, Nizhniy Novgorod, Russia).

**DNA Extraction, Polymerase Chain Reaction, and Sequencing**

HBV DNA was extracted from 5 µl of serum by proteinase K treatment as described previously [Norder et al., 1990]. Four pairs of primers were used for amplifying the S-gene: hep3-7hep5b, hep3-hep33, hep3-hep50, and hep4-hep50 [Norder et al., 1992a].

The PCR amplificates were purified using the GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences, Upplands, Sweden). Purified products were used as templates in the sequencing reaction using the dideoxynucleotide chain termination method with ABI PRISM™ BigDye™ Terminator Cycle Sequencing Reaction Kit (Applied Biosystems, a Division of Perkin-Elmer, version 3.0) with hep3, hep38, and hep50 as sequencing primers. The ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems) was used for electrophoresis and data collection.

**Sequence Analysis**

The sequences obtained were edited using the SeqMan program in the LASERGENE package (DNA STAR, Inc., Madison, WI) and aligned with the corresponding region of 156 sequences retrieved from GenBank. Phylogenetic analysis was carried out with the PHYLIP package version 3.53 [Felsenstein, 1993]. Evolutionary distances were estimated with the DNA-DIST program using the Kimura 2-parameters model. Phylogenetic trees were constructed using the UPGMA method. Bootstrap analysis was performed on 1,000 replicas with the SEQBOOT and CONSENSE programs. The HBsAg subtype of the strains was assessed from the substitutions at codons 122, 127, and 160.

**RESULTS**

It was possible to sequence the S-gene for 205 HBV strains. Three genotypes were found, D accounted for 81%, A for 18.5%, and C for 0.5% (Table I). In East Siberia (Yakutsk) there was an obvious predominance of genotype A (88%). Genotype D accounted for 86% of the isolates from Estonia and for 92–100% of the isolates from regions in the Middle Asia (Kazakhstan, Uzbekistan, and Turkmenistan) (Table I). In all 530 sera were available, 174 were from patients with acute hepatitis B, 26 from persons with chronic hepatitis B, and 330 from blood donors.

**Methods of HBsAg Detection**

Samples collected in Estonia during 1996–2002 were tested by enzyme immunoassay (EIA) with commercial kits for HBsAg detection from Organon Technika, Sanofi Diagnostics Pasteur, Murex Diagnostics, Ortho Diagnostic Systems, and DiaSorin. Sera collected in 1989, 1991, and 1997 in the other territories of the former USSR were investigated by EIA and passive agglutination assays with commercial kits (Diagnostic Systems, Nizhniy Novgorod, Russia).

**DNA Extraction, Polymerase Chain Reaction, and Sequencing**

HBV DNA was extracted from 5 µl of serum by proteinase K treatment as described previously [Norder et al., 1990]. Four pairs of primers were used for amplifying the S-gene: hep3-7hep5b, hep3-hep33, hep3-hep50, and hep4-hep50 [Norder et al., 1992a].

The PCR amplificates were purified using the GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences, Uppsala, Sweden). Purified products were used as templates in the sequencing reaction using the dideoxynucleotide chain termination method with ABI PRISM™ BigDye™ Terminator Cycle Sequencing Reaction Kit (Applied Biosystems, a Division of Perkin-Elmer, version 3.0) with hep3, hep38, and hep50 as sequencing primers. The ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems) was used for electrophoresis and data collection.

**Sequence Analysis**

The sequences obtained were edited using the SeqMan program in the LASERGENE package (DNA STAR, Inc., Madison, WI) and aligned with the corresponding region of 156 sequences retrieved from GenBank. Phylogenetic analysis was carried out with the PHYLIP package version 3.53 [Felsenstein, 1993]. Evolutionary distances were estimated with the DNA-DIST program using the Kimura 2-parameters model. Phylogenetic trees were constructed using the UPGMA method. Bootstrap analysis was performed on 1,000 replicas with the SEQBOOT and CONSENSE programs. The HBsAg subtype of the strains was assessed from the substitutions at codons 122, 127, and 160.

**RESULTS**

It was possible to sequence the S-gene for 205 HBV strains. Three genotypes were found, D accounted for 81%, A for 18.5%, and C for 0.5% (Table I). In East Siberia (Yakutsk) there was an obvious predominance of genotype A (88%). Genotype D accounted for 86% of the isolates from Estonia and for 92–100% of the isolates from regions in the Middle Asia (Kazakhstan, Uzbekistan, and Turkmenistan), West and East Siberia (Tyumen, Kemerovo and Krasnoyarsk, Chita) and the Southern part of the Far-East (Khabarovsk) (Table I). The prevalence of genotype D was lower, 50–65%, in the Central-European part of Russia (Moscow and Volga region) and 67% in the Western territories of the former USSR (Latvia, Belarus). All 58 isolates belonging to genotype A were found in one of the two major clusters of this genotype, sharing homology with mainly West-European strains. One dominant strain comprised 13 identical isolates, one from a
patient with chronic hepatitis B in Estonia, one from a patient with acute hepatitis B in Latvia, three from blood donors originating from Moscow and the Volga region, and eight from blood donors in Yakutsk (Fig. 1a). One hundred sixty one of the 166 genotype D isolates divided into two clusters although this subdivision was not supported by significant bootstrap values (Fig. 1b). By pairwise comparison, the mean nucleotide difference was found to be 0.77% (range 0.44–1.17) within the clusters, and 1.49% (range 0.73–2.79) between the two clusters. The major genotype D cluster was represented by 125 isolates almost all specifying \textit{ayw3}, 71 from patients in Estonia, eight from inhabitants of the European, and 46 from individuals in the Asian parts of the former USSR. These isolates shared similarity with strains from Germany, Poland, India, and Japan. Five dominant strains (I–V) were recognised within this cluster, represented by 47, 5, 7, 13, and 8 isolates with identical sequences (Fig. 1b). Strain I, represented by 47 isolates, comprised 41 isolates from Estonia, one each from Kemerovo and Tyumen, two isolates from blood donors in Krasnoyarsk, and two from patients in Khabarovsk. Injecting drug use was identified as source of infection for at least 15 of the 41 Estonian patients. One patient had been subject to medical interventions. Four had been in contact with HBV infected persons, while heterosexual transmission was suspected for another two patients.

Strain II was represented by four isolates from the Far-East of Russia (Khabarovsk) and one from Kazakhstan. Strain III represented by seven isolates was confined to Estonia. Medical treatment was found as risk factor for three cases and three patients gave a history of injecting drug use. A fourth strain (IV) included six isolates from Estonia and seven from West and East Siberia (Kemerovo, Krasnoyarsk). Medical intervention was the risk factor for three of the Estonian patients. One case was a member of the staff at a haemodialysis unit. The fifth dominant strain (V) was represented by one Latvian isolate from a patient with acute hepatitis B, two isolates from patients with acute hepatitis B in Khabarovsk, and five isolates from blood donors in East Siberia (Chita and Krasnoyarsk). The second genotype D cluster was represented by 36 isolates almost all specifying \textit{ayw2} originating mainly from the Central-European part of the former USSR (Moscow and Volga region), and Middle Asia (Uzbekistan, Kazakhstan and Turkmenistan). These isolates shared homology with strains from Europe, South Africa, India, Japan, and Central America. Only three isolates from Estonia belonged to this cluster. Three strains, represented by more than one isolate, were found in this cluster. One consisted of three identical isolates from the Volga region (VI), the other one was represented by three isolates from Uzbekistan (VII) and the third by two isolates from Khabarovsk (Fig. 1b).

Five strains were divergent and did not cluster with the other genotype D strains. Two of these strains derived from Estonia, two from the Volga region and one from Siberia (Fig. 1b). The single isolate of genotype C from a patient in the Far-East (Khabarovsk) shared similarity with isolates from Korea.

### TABLE I. Origin of HBsAg Positive Sera and Results From HBV PCR, Sequencing, and Subtyping

<table>
<thead>
<tr>
<th>Regions and years of collection</th>
<th>Number of studied sera</th>
<th>HBV PCR positive</th>
<th>Number of sequenced samples</th>
<th>Genotype$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>\textit{adw2}</td>
</tr>
<tr>
<td>European part of the former USSR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia, 1996–2002</td>
<td>134</td>
<td>103</td>
<td>88</td>
<td>12 (14%)</td>
</tr>
<tr>
<td>Latvia, 1991</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>2 (33%)</td>
</tr>
<tr>
<td>Belarus, 1991</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1 (33%)</td>
</tr>
<tr>
<td>Moldavia, 1991</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Moscow region, Russia, 1989, 1997</td>
<td>67</td>
<td>8</td>
<td>4</td>
<td>2 (50%)</td>
</tr>
<tr>
<td>Volga region, Russia, 1991–1997</td>
<td>82</td>
<td>28</td>
<td>17</td>
<td>6 (31%)</td>
</tr>
<tr>
<td>Asian part of the former USSR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Siberia, Russia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyumen, 1997</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Kemerovo, 1997</td>
<td>14</td>
<td>11</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Eastern Siberia, Russia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krasnoyarsk, 1997</td>
<td>37</td>
<td>17</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Chita, 1997</td>
<td>20</td>
<td>11</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Far-East, Russia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yakutsk, 1997</td>
<td>69</td>
<td>20</td>
<td>16</td>
<td>14 (88%)</td>
</tr>
<tr>
<td>Khabarovsk, 1997</td>
<td>32</td>
<td>27</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Middle Asia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kazakhstan, 1997</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Uzbekistan, 1989, 1997</td>
<td>44</td>
<td>27</td>
<td>12</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Turkmenistan, 1991</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>530</td>
<td>272 (51%)</td>
<td>205 (75%)</td>
<td>38 (18.5%)</td>
</tr>
</tbody>
</table>

$^a$One genotype C strain and one genotype D strain both from Khabarovsk specifying \textit{adr} and \textit{adw3}, and four genotype D strains with undermined subtypes, originating from Estonia, the Volga region, Kazakhstan and Chita are not included in the table.
Fig. 1. UPGMA dendrogram based on phylogenetic analysis of the 363 complete S-gene sequences representing (a) the sequences formed by 38 genotype A strains and (b) the sequences formed by 166 genotype D strains. Sequences of 156 HBV isolates from GenBank are indicated according to accession numbers. The designation and origin of each strain is indicated at the nodes of the branches. The number of identical isolates is shown within parenthesis. Roman numerals designate strains with more than one isolate.
HBV Strains in Estonia and Russia

Fig. 1. (Continued)
Within genotype D, 115 isolates belonged to subtype ayw2, 35 to ayw3, and 11 to ayw4 (Table I). One isolate from Khabarovsk encoded Lys122 and Thr127 and specified the putative subtype adw3. All strains specifying ayw4 encoded Ile127. Four other isolates from Estonia, Kazakhstan, the Volga region and Chita all with Lys160, had aberrant substitutions at subtype-specifying residues. The strains 4404-97 and 288-00 encoded Ser122, while 4386-97, 4496-97 and the mentioned 4404-97 encoded Ala127. In the major genotype D cluster, there was a subcluster represented by 100 isolates characterised by Val118 and Val129 otherwise present in strains from India. One strain, 4500-97, specifying ayw3 from a blood donor in Chita had substitutions to Leu143 and an Ala145 in the second loop of the major hydrophilic region of the a-determinant. One genotype A strain, 4588-97, from a blood donor in Yakutsk had an insertion of three amino acids (Gly-Ser-Asn) between codons 113–114 according to the enu-Yakutsk had an insertion of three amino acids (Gly-Ser-Asn) between codons 113–114 according to the enu-variant A. Out of the aberrant five strains, two specified ayw3, one strain from Estonia specified ayw4, while for two strains encoding Ser122 the subtype could not be defined.

**DISCUSSION**

Three HBV genotypes, A, D, and C were encountered in Estonia and other regions of the former USSR, with D being most prevalent. Despite genotypes B and C are prevalent in East Asia [Norder et al., 1993], only one isolate from Khabarovsk belonged to genotype C.

The majority of the genotype D isolates from Estonia and the former USSR segregated into the major cluster specifying ayw3 and was mainly represented by isolates from West-European territories (Estonia, Latvia, Belarus) on one hand, and isolates from the Siberian and Far-Eastern regions of the former USSR on the other hand. Ninety-three percent of Estonian and 80% of Siberian and Far-Eastern isolates belonged to this cluster. The second major cluster, specifying ayw2, contained strain more similar to strains from Europe, India, South Africa, and Central America, and was mainly represented by isolates from Central-European (Moscow, Volga region) and Middle Asian parts of the former USSR with relatively less isolates from Siberia and Far-East. This pattern might be explained by the ancient origin of the Estonian population from Uralic-speaking people as apparent from linguistic data [Austerlitz, 1987]. However, the explanation might also be more recent and linked to migrations in pre-revolutionary Russia or at the time of the Second World War.

For most isolates the subtypes deduced from the amino acids sequences were in agreement with the subtypes characteristic for the respective genotypes. Thus, all genotype A strains specified adw2, the single strain specifying adr° belonged to genotype C. The most prevalent subtypes in genotype D were ayw3 and ayw2 with a tendency to segregation into separate clusters. Using a panel of monoclonal antibodies Netesova et al. [2003] reported a 94% prevalence of ayw3variantA in drug users. While this subtype was found in 11% of blood donors it was absent in the native population of West Siberia. In the present study, 75% of subtype ayw3 isolates, including the major genotype D strain, represented by 47 identical isolates was related to injecting drug use had Val118 and Val129 corresponding to the ayw3variantA previously encountered in the USSR and the Balkans [Swenson et al., 2001]. These substitutions are also common in Indian strains [Norder et al., 1993; Gandhe et al., 2003]. The results of subtyping were mainly in agreement with previous reports on subtype distribution in the different regions of the former USSR [Grannikova et al., 1977]. Unexpectedly, several strains from Estonia, Siberia, and the Far-East of Russia specifying ayw4, were found to belong to genotype D. Subtype ayw4 is the only subtype specified by genotype E, and for genotype D this subtype has previously only been described for the MS-2 strain [Norder et al., 1993]. In this study, two strains, represented by seven and three isolates, and one unique Estonian isolate within genotype D encoded ayw4. All were unrelated to MS-2. The strain represented by seven identical isolates was one of the epidemiologically stable variants of HBV in Estonia. The other strain comprised two identical isolates from the Far-East and one isolate from Siberia. As for MS-2 all these strains encoded Ile127, while in genotype E strains specifying w4 encode Leu127.

Four strains from different regions of the former USSR, including Estonia, had unique substitutions not described previously at the subtype specifying residues Ser122 and Ala127. Another strain specifying ayw3 had two substitutions, Leu143 and Ala145, in the second loop of the a-determinant. A mutation Gly145Arg was described previously in vaccines in Italy [Carman et al., 1993; Gandhe et al., 2003]. The substitution Gly145Ala has been reported in an HBsAg negative variant of HBV specifying adr° in Japan [Okamoto et al., 1989; Koyanagi et al., 2000].

When comparing S-gene sequences of strains from patients in Estonia from 1996 to 2002 and from patients and blood donors collected in the former USSR in 1989–1997, a large number of identical sequences were found. Round 50% of the isolates belonged to eight strains, characterised by identical S-gene sequences, often originating from regions distantly located from each...
other and from different patient groups. For some strains there were only few nucleotide differences as compared to the dominant strains.

It was not possible in general to find a common source of infection for patients infected with any of the dominant strains, although epidemiological links were found for some Estonian patients. Thus, a number isolates from persons with a history of injecting drug use were found to belong to the major genotype D strain, while individuals with a history of medical intervention were relatively more frequently infected by the two other dominant strains. The finding of numerous identical HBV S-gene sequences of strains collected over a period of at least 13 years in distantly located regions supports the existence of genetically stable variants of HBV in the territory of the former USSR and possibly elsewhere.

ACKNOWLEDGMENTS

We thank Dr. Jelena Schmidt from Ida Viru Central Hospital for providing clinical data, Dr. Olga Lumijõe for collection of sera and Sofia Sidorova for technical assistance. The sequences reported herein are deposited in GenBank with accession numbers AY653775-AY653895.

REFERENCES