Identification of *Hepatitis B virus* putative intergenotype recombinants by using fragment typing

Jie Yang,1,2 Ke Xing,1 Riqiang Deng,1 Jinwen Wang1 and Xunzhang Wang1

1State Key Laboratory for Biocontrol, School of Life Science, Sun Yat-sen (Zhongshan) University, Guangzhou 510275, People’s Republic of China
2Department of Infectious Diseases, Nanfang Hospital, Guangzhou 510515, People’s Republic of China

Correspondence
Xunzhang Wang
wxz@zsu.edu.cn

Eight hundred and thirty-seven human *Hepatitis B virus* (HBV) genomes were categorized into pure genotypes and potential intergenotypes, according to their fragment types which were determined based on similarity and phylogenetic analyses of 13 contrived fragments of 250 bp against the corresponding fragments of the consensus sequences of genotypes A–H. Twenty-five intergenotypes, including 171 genomes, were revealed from the potential intergenotype recombinants by phylogenetic analysis of the precisely derived mosaic fragments. Among these, four new intergenotypes were discovered. Many genomes were revealed as putative intergenotype recombinants for the first time. About 87 % of the putative recombinants were B/C (120) and A/D (29) hybrids. The other recombinants comprised A/B/C, A/C, A/E, A/G, C/D, C/F, C/G, C/U (U for unknown genotype) and B/C/U hybrids. Genotypes A and C showed a higher recombination tendency than did other genotypes. The results also demonstrated region priority and breakpoint hot spots in the intergenotype recombination. Recombination breakpoints were found to be concentrated mainly in the vicinity of the DR1 region (nt 1640–1900), the pre S1/S2 region (nt 3150–100), the 3'-end of the C gene (nt 2330–2450) and the 3'-end of the S gene (nt 650–830). These results support the suggestion that intergenotype recombinants may result from co-infection with different genotypes.

**INTRODUCTION**

Human *Hepatitis B virus* (HBV), which is the prototype member of the family *Hepadnaviridae*, is a circular, partially double-stranded DNA virus of approximately 3200 bp with four overlapping ORFs encoding the polymerase (P), core (C), surface (S) and X proteins (Lee, 1997). Based on an intergroup divergence of 8% or more in the complete nucleotide sequence, HBV sequences have been classified into genotypes A–F (Okamoto et al., 1988; Norder et al., 1992, 1994; Magnus & Norder, 1995). Genotypes G and H were suggested later (Stuyver et al., 2000; Arauz-Ruiz et al., 2002). Generally, different HBV genotypes have distinct geographical distributions. For example, genotype A is prevalent in northern and central Europe and is also found in North and South America and Africa. Genotypes B and C are common in south and east Asia. Genotype D spreads worldwide, but predominates in the Mediterranean area. Genotype E is found mainly in the western part of sub-Saharan Africa, while genotype F is found in aboriginal populations of the Americas (Norder et al., 1993). Genotype G is found in USA and France (Stuyver et al., 2000). Genotype H is mainly recorded in North and South America (Arauz-Ruiz et al., 2002). It has been assumed that the HBV genotype may influence the biology of HBV and clinical disease in hosts (Mayerat et al., 1999). Genotype C induces more severe liver disease than genotype B in Asia (Bowyer et al., 1997; Kao et al., 2000), while genotype A is associated with chronic infection more frequently than genotype D in Europe (Mayerat et al., 1999).

HBV has a high rate of diversity. The mutation rate of the genome is much higher than in other DNA viruses (Okamoto et al., 1987; Georgi-Geisberger et al., 1992) because HBV replicates similarly to retroviruses through reverse transcription of an RNA intermediate, which lacks a proof-reading function (Summers & Mason, 1982). Homologous recombination between different genotypes is a source of variation. Accumulating data show that DNA recombination is frequent in HBV infections. Recombination between HBV genotypes B and C (B/C hybrid) is prevalent in south and east Asia (Morozov et al., 2000; Sugauchi et al., 2002, 2003; Luo et al., 2004). A/D hybrids have been described in Italy and South Africa (Morozov et al., 2000; Owiredu et al., 2001). Other HBV intergenotype hybrids have also been
discovered, such as C/D hybrids in China (Cui et al., 2002; Wang et al., 2005), an A/E hybrid in Cameroon (Kurbanov et al., 2005) and a C/G hybrid in Thailand (Suwannakarn et al., 2005). Intergenotype recombination may lead to different biological characteristics of the virus and a different clinical outcome in patients. For instance, recombination of genotype B with genotype C (B/C hybrid) might affect the severity of clinical disease (Sugauchi et al., 2002). Then, many questions are raised: how many kinds of intergenotypic recombinants exist in published HBV data; where are the recombination breakpoints located; and are there any rules for the recombination? These questions may lead to an understanding of the mechanism of HBV intergenotype recombination. In this paper we report 25 intergenotypes comprising 171 putative intergenotype recombinants found in 837 complete HBV genome sequences.

METHODS

HBV data. Complete human HBV genome sequences were collected and updated every month from GenBank, EMBL and DDBJ by BLASTN searches using the complete HBV genome sequences of genotypes A–H. Unique accession numbers were collected by eliminating duplicates in a text file. Sequences were then downloaded from GenBank using the Batch Entrez tool on the NCBI (National Center for Biotechnology Information) website (www.ncbi.nlm.nih.gov). After removing identical entries, patents, artificial mutants, partial sequences and animal HBV sequences, genome sequences with lengths between 3000 and 3300 bp were changed so that they started with the traditional hypothetical EcoRI site. Finally, 837 genomes consisting of 146, 152, 292, 142, 40, 42, 13 and 10 of genotypes A–H, respectively, were obtained in March 2006 (see Table 2 for a complete list of accession numbers).

Creation of consensus sequences for genotypes A–H. All the genome sequences were classified into genotypes A–H by using BLASTN (Altschul et al., 1990) and phylogenetic analysis against 24 representatives of genotypes A–H which were selected at random, three for each genotype, from HBV full-length sequences of genotypes A–H, according to Norder et al. (2004). Consensus sequences of all genotypes except genotype B were then created from all the full-length sequences of the corresponding genotypes. Genotype B consists of subgenotypes Bj (j for Japan) and Ba (a for Asia) (Sugauchi et al., 2002). Because the majority of genotype B sequences are of subgenotype Ba, which is considered to be a B/C hybrid with genotype C clustered in the preC/C region, the consensus sequence are of subgenotype Ba, which is considered to be a B/C hybrid with genotype C of all genotypes except genotype B were then created from all the sequences at a time: one putative recombinant sequence, two consensus sequences of the parental genotypes and one consensus sequence of another genotype as an outgroup. Each informative site supports one of three possible phylogenetic relationships among the four taxa. Bootscanning and cluster analysis maximizing $\chi^2$ were used to identify the breakpoints. $P$ values for the subsequent division of the sequence into genotypes were calculated by using Fisher’s exact test. After determination of recombination breakpoints, mosaic fragments were derived precisely from the recombinants, according to the breakpoints. Mosaic fragments with the same breakpoints were collected and aligned with corresponding fragments of genotype A–H representatives. Phylogenetic trees were reconstructed based on these alignments and compared with the trees reconstructed based on the correlated complete genome sequences. A 75% bootstrap value cut-off was used to identify the mosaicism of each fragment.

RESULTS

HBV genomes: classification and creation of consensus sequences

Genotypes of most HBV genomes could be determined by using BLASTN against the 24 representatives of genotypes A–H shown in Table I and by analysis of the BLASTN results based on a nucleotide sequence identity cut-off (> 92 %). Others required further phylogenetic analysis. Of the 837 HBV genomes, 146, 152, 292, 142, 40, 42, 13 and 10 were categorized in genotypes A–H, respectively. Of the 837 HBV genomes, 631 were found to have a standard genotype
Identification of HBV intergenotype recombinants

Table 1. Twenty-four full-length representatives of genotypes A–H

<table>
<thead>
<tr>
<th>Genotype</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative 1</td>
<td>M57663 (A1)</td>
<td>D50521 (B1)</td>
<td>AF411411 (C1)</td>
<td>Y07587 (D1)</td>
<td>AB091255</td>
<td>AB090458 (F1)</td>
<td>AF405706</td>
<td>AB059660</td>
</tr>
<tr>
<td>Representative 2</td>
<td>X70185 (A2)</td>
<td>AY217358 (B2)</td>
<td>AF223960 (C2)</td>
<td>AB090268 (D2)</td>
<td>X75657</td>
<td>AB036910 (F2)</td>
<td>AF160501</td>
<td>AY090454</td>
</tr>
<tr>
<td>Representative 3</td>
<td>X51970 (A3)</td>
<td>D00331 (B3)</td>
<td>X75656 (C3)</td>
<td>AB048703 (D4)</td>
<td>X75664</td>
<td>X75658 (F2)</td>
<td>AB056515</td>
<td>AY090460</td>
</tr>
<tr>
<td>Standard length</td>
<td>3221</td>
<td>3215</td>
<td>3215</td>
<td>3182</td>
<td>3212</td>
<td>3215</td>
<td>3248</td>
<td>3215</td>
</tr>
</tbody>
</table>

Table 2. Lengths of genome sequences of some representatives

<table>
<thead>
<tr>
<th>Genotype</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>(bp)</td>
<td>3221</td>
<td>3215</td>
<td>3215</td>
<td>3182</td>
<td>3212</td>
<td>3215</td>
<td>3248</td>
<td>3215</td>
</tr>
</tbody>
</table>

length, including 96, 128, 215, 105, 33, 37, 9 and 8 genomes for genotypes A–H, respectively, and these were chosen for consensus sequence creation. Consensus sequences for genotypes except genotype B were then created from alignments of genome sequences of the corresponding genotypes. For the reason explained previously, the consensus sequence of genotype B was obtained by the alignment of 28 Bj subgenotype sequences which were selected according to Sugauchi et al. (2002). The resulting consensus sequences were used as reference sequences for genotypes A–H in the subsequent fragment typing and SimPlot bootscanning analyses.

Potential intergenotype identification using fragment typing

Using BLASTN against the consensus fragments which were derived by splitting the alignment of consensus genome sequences into 13 subalignments of 250 bp, 633 genomes were identified to be pure genotypes A–H, and 204 to be potential intergenotypes, implying intergenotype recombinants. Fragment types of the 204 potential intergenotype genomes were determined further by phylogenetic analysis of potential mosaic fragments. Phylogenetic trees were reconstructed based on alignments of the potential mosaic fragments together with correlated fragments from representative genomes. Mosaic fragments that were not supported by the phylogenetic trees were modified according to the phylogenetic results, changing some originally impure fragment type genomes to pure types and others to different fragment types. The fragment types of the 837 HBV genome sequences determined with the above methods are shown in Table 2.

Identification of putative recombinants

By SimPlot analysis, the recombination breakpoints of the potential recombinants were determined precisely. Mosaic fragments were derived from the intergenotype recombinants according to the breakpoints. A phylogenetic tree was inferred to determine further the genotype of each precisely derived mosaic fragment using a 75 % bootstrap value cut-off. For example, eight genomes with the fragment type ‘ADADAAAAAAAAD’ and four genomes with the fragment type ‘DDADAAAAAAAAN’ belonged to genotype A in the whole genome, and all had a mosaic fragment between nt 209 and 526. A phylogenetic tree was then inferred, based on the alignment of these mosaic fragments and corresponding fragments of representative genomes of genotypes A–H. All the 12 mosaic fragments clustered with the corresponding representative fragments of genotype D with a bootstrap value of 93 % in the tree (Fig. 1a). A similar tree was also inferred based on alignment of the corresponding whole genomes. All 12 genome sequences clustered with the representative genomes of genotype A with a bootstrap value of 100 % (Fig. 1b). Phylogenetic trees based on other mosaic fragments were also reconstructed to demonstrate the mosaicism using a 75 % bootstrap value cut-off (data not shown). The mosaicism of the precisely derived fragments had much stronger bootstrap support than did that of the contrived fragments.

Finally, 166 genomes were revealed as putative intergenotype recombinants supported by phylogenetic trees. According to the fragment types, these genomes were classified into 23 intergenotypes. For convenience of discussion, the 23 intergenotypes were named intergenotype 1–23 (IG-1 to IG-23 as indicated in Table 2 and Table 3). Fig. 2 shows the SimPlot bootscanning results and Table 3 lists the breakpoints of representative genome sequences of the 23 intergenotypes. The bootstrap values for each mosaic fragment are also listed in Table 3. Sequences in the same intergenotype had similar breakpoints. The 5′ terminal breakpoints and the 3′ terminal breakpoints of the 117 genomes in IG-6 were located within nt 1730–1846 and 2244–2485 respectively. Other intergenotypes had breakpoints with variations of only tens of base pairs.

In addition to the above 166 genomes, constituting the 23 intergenotypes, there are five genomes that need further investigation and are currently categorized to two uncertain intergenotypes. Genomes AF241407–AF241409 and AB231908 were originally from Vietnam. These four genomes exhibited a complicated fragment type of ‘NAGGGECCCCCCCN’ in the fragment typing stage, where N represents an uncertain genotype of the 250 bp contrived fragment, i.e. the fragment did not cluster with any of the representative fragments of the eight genotypes in the tree. In the following analysis stage with precisely derived mosaic fragments, the mosaic areas corresponding to A, G and E were not supported by bootstrap values higher than the 75 % cut-off. SimPlot analysis also showed that the corresponding area between nt 2866 and 1800 was significantly different.
Table 2. Fragment type analysis of the 837 HBV complete genome sequences

Accession numbers and countries are listed alphabetically according to the fragment types. Continuous accession numbers were combined for parity: for instance, ‘AB116076–94’ stands for 19 genomes from AB116076 to AB116094. The latterly identified 25 intergenotypes are indicated in the second column (IG, intergenotype). C/Gi, Genotype C and gibbon hybrid suggested by Simmonds & Midgley (2005).

<table>
<thead>
<tr>
<th>Fragment type</th>
<th>IG</th>
<th>Genotype</th>
<th>No.</th>
<th>Country</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBBBBBBBBBBB B</td>
<td>32</td>
<td>Japan</td>
<td>AB010289–92, AB014366, AB073838, AB073842–58, AB106884–85, AB205121, D00329, D23677–79, D50521–22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFFFFFFFFFFF F</td>
<td>41</td>
<td>Argentina, Bolivia, Costa Rica, El Salvador, Japan, Germany, Panama, Nicaragua, Sweden, USA, Venezuela</td>
<td>AB036905–20, AB064316, AB086397, AB116549–52, AB116654, AB166850, AF223962–65, AY090455–56, AY090458, AY090459, AY090461, AY179734–35, AY311369–70, NT_033927, X69798, X75658, X75663</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. cont.

<table>
<thead>
<tr>
<th>Fragment type</th>
<th>IG</th>
<th>Genotype</th>
<th>No.</th>
<th>Country</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGGGGGGGGGGG</td>
<td>G</td>
<td>11</td>
<td>Belgium, Japan, Germany, USA</td>
<td>AB056513–15, AB064310–13, AF160501, AF405706, AP007264, DQ207798</td>
<td></td>
</tr>
<tr>
<td>HHHHHHHHHHHHH</td>
<td>H</td>
<td>10</td>
<td>Japan, Nicaragua, USA</td>
<td>AB059659, AB059660, AB059661, AB064315, AB179747, AB205010, AP007261, AY090454, AY090457, AY090460</td>
<td></td>
</tr>
<tr>
<td>ADADAAAAAADA</td>
<td>IG-1</td>
<td>A</td>
<td>8</td>
<td>India</td>
<td>AF418684–89, AY161148–49</td>
</tr>
<tr>
<td>DDADAAAAAAN</td>
<td>IG-2</td>
<td>A</td>
<td>4</td>
<td>India</td>
<td>AF418690–92, AY161147</td>
</tr>
<tr>
<td>AAADADDDNAAA</td>
<td>IG-3</td>
<td>A</td>
<td>8</td>
<td>India</td>
<td>AF418674–75, AY161140–41, AY161145–46, AF418682–83</td>
</tr>
<tr>
<td>DDAAAAAANAND</td>
<td>IG-4</td>
<td>A</td>
<td>2</td>
<td>South Africa</td>
<td>AF297619, AF297620</td>
</tr>
<tr>
<td>AAADADADAAN</td>
<td>IG-5</td>
<td>A</td>
<td>1</td>
<td>South Africa</td>
<td>AF297621</td>
</tr>
<tr>
<td>BBBBBBBCNBBBA</td>
<td>IG-7</td>
<td>B</td>
<td>1</td>
<td>South Africa</td>
<td>U87747</td>
</tr>
<tr>
<td>CCCCCCCCCCCCC</td>
<td>IG-8</td>
<td>C</td>
<td>1</td>
<td>China</td>
<td>AY057947</td>
</tr>
<tr>
<td>CCCCCCCBCCCCC</td>
<td>IG-9</td>
<td>C</td>
<td>1</td>
<td>Vietnam</td>
<td>AB031265</td>
</tr>
<tr>
<td>CCCCCCCBCCCCC</td>
<td>IG-10</td>
<td>C</td>
<td>1</td>
<td>China</td>
<td>AF233236</td>
</tr>
<tr>
<td>CCCCCCCBCCCCC</td>
<td>IG-11</td>
<td>C</td>
<td>1</td>
<td>Japan</td>
<td>D16665</td>
</tr>
<tr>
<td>DDDDDCCDDCCDD</td>
<td>IG-12</td>
<td>C</td>
<td>8</td>
<td>China</td>
<td>AF461043, AY800249, AY817509–10, AY8175–15</td>
</tr>
<tr>
<td>DDDDDDCDDCCDD</td>
<td>IG-13</td>
<td>C</td>
<td>2</td>
<td>China</td>
<td>AY057948, AY187511</td>
</tr>
<tr>
<td>DDAADDDDDDDDD</td>
<td>IG-14</td>
<td>D</td>
<td>1</td>
<td>Italy</td>
<td>AY236161</td>
</tr>
<tr>
<td>DDAADDDADDDDD</td>
<td>IG-15</td>
<td>D</td>
<td>1</td>
<td>Italy</td>
<td>X65258</td>
</tr>
<tr>
<td>DDAADDDADDDDD</td>
<td>IG-16</td>
<td>D</td>
<td>1</td>
<td>Italy</td>
<td>X65259</td>
</tr>
<tr>
<td>DDAADDDADDDDD</td>
<td>IG-17</td>
<td>D</td>
<td>1</td>
<td>Italy</td>
<td>X68292</td>
</tr>
<tr>
<td>DDAADDDADDDDD</td>
<td>IG-18</td>
<td>D</td>
<td>2</td>
<td>India</td>
<td>AF418681, AY161161</td>
</tr>
<tr>
<td>AAADAAEAAAAA</td>
<td>IG-19</td>
<td>A</td>
<td>1</td>
<td>Cameroon</td>
<td>AB194949</td>
</tr>
<tr>
<td>FFFFFFFFFFFF</td>
<td>IG-20</td>
<td>F</td>
<td>1</td>
<td>Bolivia</td>
<td>AB214516</td>
</tr>
<tr>
<td>AGGGGGGGGGGGG</td>
<td>IG-21</td>
<td>G</td>
<td>1</td>
<td>USA</td>
<td>AB056516</td>
</tr>
<tr>
<td>GGGGGGGGGGGGG</td>
<td>IG-22</td>
<td>G</td>
<td>1</td>
<td>Thailand</td>
<td>DQ078791</td>
</tr>
<tr>
<td>BBBBAACCNBBB</td>
<td>IG-23</td>
<td>B</td>
<td>1</td>
<td>Philippines</td>
<td>AB219430</td>
</tr>
<tr>
<td>NAGGGECCCCCN</td>
<td>IG-24</td>
<td>C</td>
<td>4</td>
<td>Vietnam</td>
<td>AB231908, AF241407–09</td>
</tr>
<tr>
<td>BBGGGECNNBBB</td>
<td>IG-25</td>
<td>B</td>
<td>1</td>
<td>Vietnam</td>
<td>AB231909</td>
</tr>
<tr>
<td>CNNCCCDCCCDD</td>
<td>C/Gi</td>
<td>C</td>
<td>2</td>
<td>Australia</td>
<td>AB048704–05</td>
</tr>
<tr>
<td>DDDDDDDDDDDDD</td>
<td>D</td>
<td>3</td>
<td>Australia</td>
<td>AB048701–03</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>837</strong></td>
</tr>
</tbody>
</table>
from all known genotypes. Hannoun et al. (2000) had supposed that fragments like these were from an unknown genotype (U) or a new genotype. We considered the four genomes as recombinants between genotype C and an unknown genotype (C/U hybrid) and assigned another intergenotype (IG-24) to include these four recombinants. Another genome (AB231909) with fragment type ‘BBBGG-ECCCNBBB’ belonged to subgenotype Ba in the main part of the genome. SimPlot analysis showed that the fragment corresponding to the fragment type GGEC area between nt 729 and 1895 showed a high level of identity with the unknown genotype fragment in IG-24 (99 %). This genome was considered to be a Ba/U hybrid or a B/C/U hybrid and was assigned to intergenotype IG-25 (Tables 1 and 2, and Fig. 2).

**Putative intergenotype recombinants**

Among the 25 intergenotypes, four were identified for the first time (marked with ‘*’ in Table 3). In the other 21 intergenotypes already reported (see references in Table 3) many genomes were also revealed for the first time.

**A/D hybrids.** Ten intergenotypes, including 29 HBV genomes, were revealed to be A/D hybrids. Five intergenotypes (IG-1 to IG-5) belonged to genotype A, including 20 genomes from India and 3 from South Africa. IG-14 to IG-18, including six genomes (four from Italy and two from India), belonged to genotype D. A/D hybrids from India mostly belonged to genotype A in the whole genome, while those from Italy all belonged to genotype D. Putative mosaic areas of A/D hybrids were located mainly in the long S and X/preC gene region.

**B/C hybrids.** One hundred and seventeen genomes in IG-6 belonged to subgenotype Ba and the mosaic area just covered the preC/C region (Sugauchi et al., 2002). B/C hybrids composed about 79% (120 out of 152) of all genotype B sequences. Genomes of pure fragment type BBBBBBBBBBBBB were found only in Japan. U87747 (IG-7) from South Africa and AB219430 (IG-23) from the Philippines were two different A/B/C hybrids. Both of them were actually hybrids of Ba recombined with a genotype A fragment at a different position: U87747 in the preS1/S2 region and AB219430 in the X gene region.

---

**Fig. 1.** Phylogenetic analysis of mosaic fragments within nt 209–526 derived from genomes with fragment types ‘ADADAAAAAAAAA’ (IG-1) and ‘DDADAAAAAAAU’ (IG-2). GenBank accession numbers are shown on each tree. The letter before each accession number indicates the genotype: upper-case letters for reference sequences of genotypes A–H and lower-case letters for putative recombinants. Upper-case letters following accession numbers indicate putative hybrids. Numbers at nodes indicate the percentage of bootstrap replications supporting the clusters (values of 75 % and higher are shown). (a) Phylogenetic tree (1000 sets) inferred from mosaic fragments within nt 209–526. (b) Phylogenetic tree (1000 sets) reconstructed based on the correlated complete genome sequences.
Table 3. Identification of mosaic fragments derived from representatives of each intergenotype

Intergenotypes marked with ‘*’ are revealed for the first time in this report, and references are listed for those reported before. G1 and G2 refer to the original and alternate parent genotypes of each intergenotype, respectively; consensus sequences were used as references. Mosaic fragments of each intergenotype representative were derived precisely from the recombinants according to the breakpoints (column ‘Fragments’). G1 (nt) and G2 (nt) refer to the number of nucleotides within each mosaic fragment which differ from G1 and G2, respectively. The similarity between each mosaic fragment and the correlated fragments from G1 or G2 was determined as a percentage (columns ‘G1’ and ‘G2’ under identity, respectively), and differences in similarity are also indicated (column ‘G2–G1’). Each mosaic fragment was characterized by a bootstrap tree and the bootstrap values are listed.

<table>
<thead>
<tr>
<th>IG</th>
<th>No.</th>
<th>Hybrid</th>
<th>G1</th>
<th>G2</th>
<th>Fragment</th>
<th>Length (nt)</th>
<th>G1 (nt)</th>
<th>G2 (nt)</th>
<th>Identity (%) G1</th>
<th>G2–G1 (%)</th>
<th>Bootstrap (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>D</td>
<td></td>
<td>781–1095</td>
<td>272</td>
<td>40</td>
<td>11</td>
<td>87-3</td>
<td>9-2</td>
<td>99-0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>D</td>
<td></td>
<td>781–1095</td>
<td>272</td>
<td>40</td>
<td>11</td>
<td>82-3</td>
<td>10-7</td>
<td>99-0</td>
</tr>
<tr>
<td>IG-4</td>
<td>2</td>
<td>A/D</td>
<td>A</td>
<td>D</td>
<td>AF297619</td>
<td>2856–417</td>
<td>783</td>
<td>111</td>
<td>27</td>
<td>96-2</td>
<td>10-6</td>
<td>99-0</td>
</tr>
<tr>
<td>IG-5</td>
<td>1</td>
<td>A/D</td>
<td>A</td>
<td>D</td>
<td>AF297621</td>
<td>1360–1604</td>
<td>245</td>
<td>18</td>
<td>6</td>
<td>92-7</td>
<td>4-9</td>
<td>93-2</td>
</tr>
<tr>
<td>IG-6</td>
<td>117</td>
<td>B/C</td>
<td>B</td>
<td>C</td>
<td>AF121243</td>
<td>1740–2443</td>
<td>704</td>
<td>55</td>
<td>18</td>
<td>89-2</td>
<td>3-6</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-7*</td>
<td>1</td>
<td>A/B/C</td>
<td>A</td>
<td>B</td>
<td>U87747</td>
<td>1740–2443</td>
<td>704</td>
<td>56</td>
<td>21</td>
<td>89-4</td>
<td>3-6</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-8</td>
<td>1</td>
<td>A/C</td>
<td>A</td>
<td>B</td>
<td>2965–3215</td>
<td>251</td>
<td>35</td>
<td>14</td>
<td>86-1</td>
<td>94-5</td>
<td>8-3</td>
<td>98-9</td>
</tr>
<tr>
<td>IG-9</td>
<td>1</td>
<td>B/C</td>
<td>B</td>
<td>A</td>
<td>AB031265</td>
<td>1219–1645</td>
<td>427</td>
<td>34</td>
<td>10</td>
<td>97-8</td>
<td>5-7</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-10</td>
<td>1</td>
<td>B/C</td>
<td>B</td>
<td>A</td>
<td>AB233256</td>
<td>1244–1799</td>
<td>556</td>
<td>44</td>
<td>11</td>
<td>92-7</td>
<td>5-9</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-11</td>
<td>1</td>
<td>B/C</td>
<td>B</td>
<td>D</td>
<td>D16665</td>
<td>1741–2443</td>
<td>695</td>
<td>54</td>
<td>13</td>
<td>92-0</td>
<td>5-9</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-12</td>
<td>8</td>
<td>C/D</td>
<td>C</td>
<td>D</td>
<td>AY187151</td>
<td>10–1484</td>
<td>1475</td>
<td>118</td>
<td>34</td>
<td>97-8</td>
<td>5-7</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-13</td>
<td>2</td>
<td>C/D</td>
<td>C</td>
<td>D</td>
<td>AY057948</td>
<td>10–1484</td>
<td>1475</td>
<td>118</td>
<td>34</td>
<td>97-8</td>
<td>5-7</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-14</td>
<td>1</td>
<td>A/D</td>
<td>A</td>
<td>Y</td>
<td>AY236161</td>
<td>400–1854</td>
<td>1455</td>
<td>110</td>
<td>69</td>
<td>92-4</td>
<td>2-8</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-15</td>
<td>1</td>
<td>A/D</td>
<td>A</td>
<td>X</td>
<td>X65258</td>
<td>422–756</td>
<td>335</td>
<td>21</td>
<td>9</td>
<td>93-7</td>
<td>3-6</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-16</td>
<td>1</td>
<td>A/D</td>
<td>A</td>
<td>X</td>
<td>X65259</td>
<td>822–1837</td>
<td>1016</td>
<td>82</td>
<td>42</td>
<td>93-6</td>
<td>3-9</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-17</td>
<td>1</td>
<td>A/D</td>
<td>A</td>
<td>X</td>
<td>X68292</td>
<td>680–1219</td>
<td>539</td>
<td>44</td>
<td>22</td>
<td>95-5</td>
<td>4-3</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-19</td>
<td>1</td>
<td>A/E</td>
<td>A</td>
<td>E</td>
<td>AB194949</td>
<td>874–1073</td>
<td>200</td>
<td>25</td>
<td>5</td>
<td>87-6</td>
<td>3-6</td>
<td>99-5</td>
</tr>
<tr>
<td>IG-20</td>
<td>1</td>
<td>C/F</td>
<td>F</td>
<td>C</td>
<td>AB214516</td>
<td>1558–1844</td>
<td>287</td>
<td>29</td>
<td>11</td>
<td>89-9</td>
<td>6-3</td>
<td>98-0</td>
</tr>
<tr>
<td>IG-21</td>
<td>1</td>
<td>A/G</td>
<td>G</td>
<td>A</td>
<td>AB056516</td>
<td>17–217</td>
<td>201</td>
<td>17</td>
<td>7</td>
<td>91-5</td>
<td>7-5</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-22</td>
<td>1</td>
<td>C/G</td>
<td>C</td>
<td>D</td>
<td>DQ078791</td>
<td>1847–2400</td>
<td>554</td>
<td>57</td>
<td>9</td>
<td>97-4</td>
<td>8-7</td>
<td>100-0</td>
</tr>
<tr>
<td>IG-23*</td>
<td>1</td>
<td>A/B/C</td>
<td>B</td>
<td>A</td>
<td>AB219430</td>
<td>1185–1784</td>
<td>599</td>
<td>38</td>
<td>18</td>
<td>96-8</td>
<td>3-3</td>
<td>90-0</td>
</tr>
<tr>
<td>IG-24</td>
<td>4</td>
<td>C/U</td>
<td>C</td>
<td>U</td>
<td>AF241407</td>
<td>2866–1800</td>
<td>2149</td>
<td>153</td>
<td>6</td>
<td>92-9</td>
<td>6-8</td>
<td>100-0</td>
</tr>
</tbody>
</table>
AB231909 (IG-25, B/C/U hybrid) was also a hybrid of Ba recombined with an unknown genotype. The recombination breakpoints were just at the end of the S gene and at the beginning of the C gene.

Genotypes B and C are the two most prevalent genotypes in south and east Asia, but only three B/C hybrids (IG-9 to IG-11) were found from all 292 genotype C genomes: D16665 from Japan, AF233236 from China and AB031265 from Vietnam. Genome D16665 with fragment type CCCC-CCCBBCCCCC was a B/C hybrid with genotype B clustered in the preC/C region. The mosaic area of genomes AB031265 and AF233236 was at the end of the P gene and a part of the X gene.

C/D hybrids. IG-12 and IG-13, including 10 HBV isolates, were C/D hybrids. They belonged to genotype C in the whole genome. All 10 sequences are from north-west China, as well as Tibet, where genotype D is rare and where the dominant strain of HBV is a C/D hybrid (Cui et al., 2002).

A/C hybrids. Genome AY057947, the only A/C hybrid identified in this study, belongs to genotype C. The mosaic area was at the preS1/S2 region. This genome has been found in China where genotype A is seldom recorded, only one genome being recorded as genotype A in the 127 genomes from China.

Other hybrids. Genome AB214516 from Bolivia was a putative C/F hybrid with a mosaic region at nt 1535–1866, including part of the X gene and the preC region.

AB194949 from Cameroon is the only A/E hybrid, recently revealed by Kurbanov et al. (2005). The A/G hybrid genome AB056516 belonged to genotype G with a 200 bp genotype A mosaic fragment in the preS1/S2 region. A genome from Thailand, DQ078791, belonging to genotype G was also identified as a C/G hybrid. The mosaic area just covered the preC/C region (Suwannakarn et al., 2005). Two intergenotypes (IG-24 and IG-25) comprised mosaic areas of an unknown genotype.

To avoid missing some recombinants, we repeated the above analysis using the same methods, but with a change at the fragment typing stage, i.e. splitting the 13 contrived fragments in the middle, i.e. the first fragment started at nt 126 and ended at nt 375 downstream of the traditional hypothetical EcoRI site. Although this yielded 238 potential intergenotype recombinants at the fragment typing stage, subsequent phylogenetic analysis of the precisely derived mosaic fragments identified the same 171 intergenotype recombinants (data not shown).

**DISCUSSION**

Many HBV intergenotype recombinants have been revealed since the first report by Bollyky et al. (1996). Morozov et al. (2000) discovered six B/C hybrids (belonging to IG-6 or IG-11 in our study) and three A/D hybrids (IG-15 to IG-17) from 99 complete HBV sequences. Almost at the same time, Bowyer & Sim (2000) found three A/D hybrids and six B/C hybrids (IG-6) from 65 genome sequences. Several new hybrid types have been revealed subsequently. Recently,
Simmonds & Midgley (2005) revealed 24 phylogenetically
independent intergenotypes.

This study scanned 837 human HBV genomes currently
available in GenBank for intergenotype recombinants. It was
difficult to deal with such a large amount of data using
described methods (Morozov et al., 2000; Bowyer & Sim,
2000; Simmonds & Midgley, 2005), so a different strategy
was used in this study to identify potential recombinants
from these genomes (i) based on similarity analysis of 13
contrived fragments of 250 bp, and (ii) based on the pre-
cisely derived mosaic fragments. Results from the first step
indicated that most genomes were of pure genotypes, only
204 genomes being potential intergenotype recombinants.
Analysis at the second step by a phylogenetic method based
on the precisely derived mosaic fragments revealed 171
putative recombinants, which were categorized into 25
intergenotypes according to their fragment types. Among
the 25 intergenotypes, IG-6 contained 117 genomes; the
other 24 intergenotypes contained 54 genomes.

Compared to the results reported by Simmonds & Midgley
(2005), our study revealed five more intergenotypes, four
(IG-2, IG-7, IG-23 and IG-25) of which are reported here for
the first time, and IG-22 (G/C hybrid) reported recently by
Suwannakarn et al. (2005). Excluding one intergenotype
which was a recombinant of human and gibbon HBV, and
two intergenotypes of chimpanzee HBV, the remaining 21
human HBV intergenotypes reported by Simmonds &
Midgley (2005) are included in 20 of our 25 intergenotypes.
In addition, our study has revealed many more human HBV
genomes as intergenotype recombinants than did previous
studies.

Among the intergenotypes presented in this study, IG-7 and
IG-23 were two different A/B/C hybrids between subgeno-
type Ba and genotype A with different recombination posi-
tions, and IG-25 was a B/C/U hybrid between subgenotype
Ba and an unknown genotype. Though almost identical to
IG-1 in its main genome, IG-2 was a different intergenotype
due to a 183 bp deletion in the preS1/S2 region and a longer
mosaic area of genotype D in both flanks of the deletion as
compared to IG-1. It should be mentioned that AF418684–
AF418689 in IG-1 and AF418690–AF418692 in IG-2 are
sequences of different clones from the same patient (patient
313) in India. Genomes AY161148–AY161149 in IG-1 and
AY161147 in IG-2 are also from the same patient in India
(according to GenBank). Genomes of different intergeno-
types existing in serum samples from the same patient might
help in understanding the mechanism of recombination.

This study scanned all the currently available human HBV
genomes, and categorized all putative recombinants into
different intergenotypes by fragment typing, presenting
comprehensive information on human HBV intergenotypes
and including a particular description for each intergeno-
type. Hence, a comparatively complete inspection of human
HBV for intergenotype recombinants was completed, which
should be helpful for finding hot spots of recombination, for
investigating the geographical distribution of intergenotype
recombinants and for understanding the tendency of different
genotypes to recombine and the mechanism of inter-
genotype recombination in HBV.

**Hot spots of recombination breakpoints**

Different regions in the HBV genome showed different rates
of recombination, and recombination breakpoints were
found to be located frequently near gene boundaries
(Simmonds & Midgley, 2005), like at the start of the S1,
S2, X, C and P genes, and at the end of the S, X, P and C
genes. Of the breakpoints listed in Table 3, 54 or 66 % are
located within 60 or 80 bp of one of these boundaries.
Comparison between our data on intergenotype recombi-
nation and previous data revealed some hot spots for
recombination breakpoints.

Sixty percent of the intergenotypes (15 out of 25), including
A/D (4), A/B/C (2), A/E (1), B/C (4), C/F (1), C/G (1), C/U
(1) and B/C/U (1) hybrids, contained breakpoints within nt
1640–1900, which lies in the vicinity of the DR1 region. It
was shown that the recombination site density in the region
covering nt 1600–2000 was almost fivefold higher than that
of the remainder of the genome (Pineau et al., 1998). DR1
has been demonstrated to cover a hot spot (nt 1885–1915)
for genomic recombination and has been proposed as a
candidate site for intragenic recombination between HBV
isolates of different genotypes (Hino et al., 1991; Sugauchi
et al., 2002).

The second hot spot of recombination is located in the
preS1/S2 region, covering nt 3150–100. Seven intergeno-
types, including A/D (2), C/D (2), A/B/C (1), A/C (1) and
A/G (1) hybrids, showed breakpoints in this region. Results
of previous analyses of HBV partial sequences also sup-
ported this hot spot. Kato et al. (2002b) found that some
isolates in genotype G had a mosaic area of genotype A in
the preS1/S2 region. Chen et al. (2004, 2006) discovered
that some HBV isolates of genotype B were integrated with
genotype C sequences from patients co-infected with geno-
types B and C. The recombination breakpoints were also
mainly found in the preS1/S2 region.

The third recombination hot spot is located near the 3’ end
of the C gene, covering nt 2330–2485. Recombinants with
breakpoints at this spot usually have associated breakpoints
at the previous DR1 hot spot. About one-third (8 out of 25)
of the intergenotypes, including A/D (2), A/B/C (2), B/C (3)
and C/G (1) hybrids, were found to have breakpoints at this
spot. Another region between nt 650 and 830 near the 3’ end
of the S gene was also a hot spot for recombination. Nine
intergenotypes, including A/D (6), C/D (1), A/E (1) and
B/C/U (1) hybrids, had breakpoints in this area. It was
interesting that some of the A/D hybrids had breakpoints at
the 6 bp insertion of genotype A (nt 2356–2361, IG-3) or at
the 33 bp deletion of genotype D (nt 2855–2887, IG-2 and
IG-4). This reminded us of some intergenotypes with in-
frame deletions in certain areas, especially in the preS1/S2
region, like IG-2, IG-4, IG-5, IG-21 and IG-11. The relationship between deletion and intergenotype recombination requires further investigation.

Recombination tendency of different genotypes

Genotypes of the 171 putative recombinants comprised genotypes A (24 genomes), B (120 genomes), C (18 genomes), D (6 genomes), F (1 genome) and G (2 genomes). Recombinants in genotypes E and H were not found in this study. Most of the putative recombinants were B/C (120) and A/D (29) hybrids. The other recombinants comprised A/B/C (2), A/C (1), A/E (1), A/G (1), C/D (10), C/F (1), C/G (1), C/U (4) and B/C/U (1) hybrids. Genotype A was found to recombine with genotypes B, C, D, E and G, and genotype C was found to recombine with genotypes A, B, D, F, G and U. However, the other genotypes were found to recombine only with genotypes A or C. It seems that genotypes A and C have a higher recombination tendency than the other genotypes do. Recently, genotype C was even found to integrate into the genome of a chimpanzee HBV genome (Magiorkinis et al., 2005), while a human HBV genotype C isolate was supposed to have a gibbon HBV mosaic fragment (Suwannakarn et al., 2005).

It is interesting that the partial sequence of one genotype tends to integrate into other genotypes in certain regions. Genotype C tends to recombine into other genotypes in the preC/C (often covering part or all of the C gene) or the preS1/S2 region, like IG-6, IG-7, IG-22, IG-23 to IG-25 in the preC/C region, and the recombinants revealed by Chen et al. (2004, 2006) in the preC/C or preS1/S2 regions. Genotype A tends to integrate into other genotypes in X/preC (often covering part or all of the X gene) or the preS1/S2 region, like IG-14 to IG-17 and IG-23 in the X/preC region, and IG-7, IG-8 and IG-21 in the preS1/S2 region. Kato et al. (2002b) also recorded the integration of genotype A into genotype G in the X/preC and preS1/S2 regions.

Geographical distribution of putative recombinants

The results showed that, in general, intergenotypes tend to have the same geographical distributions as that of their parent genotypes. A/D hybrids in genotype A are mainly recorded in India (IG-1 to IG-3) and South Africa (IG-4), while A/D hybrids in genotype D are only found in Italy (IG-14 to IG-17) and India (IG-18). B/C hybrids in genotype C (IG-9 to IG-11) can be found in China, Japan and Vietnam, but they are only sporadic examples. Other recombinants, including A/E (IG-19) in Cameroon, A/B/C (IG-7) in South Africa, A/C (IG-8) in China, C/F (IG-20) in Bolivia, A/G (IG-21) in America and C/G (IG-22) in Thailand, only occur sporadically.

However, the B/C hybrid in genotype B (IG-6) is one of the dominant subgenotypes in south and east Asia, but the prototype of genotype B has not been found, except in Japan (Sugauchi et al., 2002). C/D hybrids (IG-12 to IG-13) are only found in the north-west of China, as well as Tibet, where C/D hybrids even become the dominant HBV groups (Cui et al., 2002), but genotype D as a parent genotype is rarely found there.

Co-infection and recombination

How intergenotype recombination occurs still remains unanswered. In general, recombination is supposed to result from co-infection of different genotypes (Fares & Holmes, 2002; Hannoun et al., 2002), though there is little chance for two HBV genomes to be recombined due to the encapsidation of only one pregenome for replication (Kato et al., 2002b). However, some cases of co-infection of different genotypes have been reported, such as co-infection of genotypes A and D (Bahn et al., 1997; Gerner et al., 1998), A and G (Kato et al., 2002a), and B and C (Chen et al., 2004, 2006). Intergenotype recombinations have been found to occur in co-infected sera, but in the cases reported so far, the sequences were only partial genome sequences. Our data have revealed a case with a complete genome sequence, an A/G hybrid found in sera co-infected by genotypes A and G.

Genotype G was found to frequently co-infect with other genotypes. All the isolates of genotype G discovered in San Francisco co-infected with genotype A (Kato et al., 2002a), while those from Canada co-infected with either genotype A or C (Osiowy & Giles, 2003). Genome AB056516 from San Francisco, an A/G hybrid in genotype G (IG-21), was proved to be derived from sera co-infected with isolates of genotypes G and A (Kato et al., 2002a). Therefore, it is suspected that co-infection caused the formation of AB056516.

Some intergenotypes of putative recombinants become prevalent in certain areas, like B/C hybrids of genotype B (IG-6) in Asia (Sugauchi et al., 2002) and C/D hybrids of genotype C (IG-13 to IG-14) in north-west China (Cui et al., 2002; Wang et al., 2005), where one of the parent genotypes rarely exists. For these hybrids, recombination might have occurred a long time ago and co-infection of the parent genotypes is no longer found in those areas. But for other hybrids, recombination might just occur in patients co-infected with more than one genotype. Co-infections were found to occur frequently in certain patients and cases like these could be very helpful for understanding the mechanism of recombination.

Intergenotype recombination is a frequent affair in HBV infection, especially in cases of co-infection. Further study is required to clarify the mechanism of recombination, the role of recombination in the heterogeneity of HBV and the outcome of recombinants, which might help in our understanding of HBV history and the variety of host responses.

ACKNOWLEDGEMENTS

This work was funded by the Natural Science Foundation of Guangdong Province under Grant 04009785.
REFERENCES


Sugauchi, F., Orito, E., Ichida, T. & 10 other authors (2003). Epidemiologic and virologic characteristics of hepatitis B virus genotype B having the recombination with genotype C. *Gastroenterology* 124, 925–932.


