The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid

Chaoyin Cui,1 Jinxiu Shi,2 Lijian Hui,2 Huifeng Xi,3 Zhuoma,1 Quni,1 Tsedan1 and Gengxi Hu2

1 Research Center of Fundamental Physics, Medical School of Tibet University, Lhasa, Tibet
2 Shanghai Life Science Center, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, 320 Yue-yang Road, Shanghai 200031, China
3 Institute of Genetics, Fudan University, Shanghai, China

There are no reports on DNA sequences of hepatitis B virus (HBV) strains from Tibet, although this highland area has a high HBsAg-positive population. We characterized HBV isolates from sera of 26 HBsAg-positive Tibetans. To determine the HBV genotypes and their phylogenetic relationships, we sequenced two genomic regions, one including the pre-S1/pre-S2/S region and the other including the pre-C/C region. The sequences were classified into two different genotypes based on different regions of the genome, except for one isolate. To clarify this finding, two complete HBV genomes that represented the two groups of isolates were sequenced. From the sequencing results, we concluded that HBV strains in Tibet may be classified as genotype C, and there are at least two subgroups. The dominant subgroup is a C/D hybrid with serotype ayw2, and the other is genotype C with serotype adw. This is the first report of complete nucleotide sequences of HBV from Tibet. These results contribute to the investigation of recombinant HBV strains throughout the world and should encourage further study of genotypes and recombination in HBV from this particular region.

Introduction

Despite the availability of HBsAg vaccines and the mass immunization schemes implemented in over 80 countries worldwide, hepatitis B virus (HBV) infection continues to be a major public health problem. There are over 350 million chronic carriers worldwide, of whom approximately 10% may die of secondary cirrhosis and hepatocellular carcinoma (Kane, 1996). Studies in molecular epidemiology over the past two decades have classified HBV DNA sequences into seven genotypes, A–G (Okamoto et al., 1986; Norder et al., 1992, 1994; Stuyver et al., 2000). The definition of the HBV genotype is based on one of the following criteria: an inter-group divergence of 8% or greater in the complete genome nucleotide sequence, or a 4–1% divergence or greater of the surface antigen gene (Okamoto et al., 1988; Norder et al., 1994). Genotypes are distributed geographically. Genotype A is found worldwide; genotypes B and C prevail in Asia; genotype D is found in southern Europe, the Americas and Australia; genotype E is most commonly found in Africa; genotype F is found in native Americans and Polynesians; and genotype G is found in the USA and Europe (Stuyver et al., 2001). The prevalent HBV strains in China are genotypes B and C (Zhu et al., 1999). HBVs with recombinant genotypes have been observed. Phylogenetic analysis has shown that B/C recombinants have spread through East Asia and that A/D recombinants exist in Italy (Morozov et al., 2000).

In Tibet, 26·2% of the population is HBsAg-positive (Luo et al., 1993). However, no HBV strain has been studied in Tibet. Undiscovered genotypes may have played a role in the high HBV infection rate. By sequencing and phylogenetic analysis of the local HBV isolates, we report here on the dominant HBV genotype in Tibet — a C/D hybrid.

Methods

Study subjects. More than 1000 serum samples positive for HBsAg were collected from primary school students with chronic asymptomatic HBV infection in Lhasa, Shigatse, Nyingchi and Tsedang.
Table 1. Primers used for HBV DNA amplification and sequencing

| Primers used for amplification | Fragment A | | Fragment B | | Fragment I | | Fragment II | | Fragment III |
|---|---|---|---|---|---|---|---|---|
| CF1: 1719-21f | GTGTTAAGGACTGGGAGGAG | SF: 2798-21f | AACACATAGCCCTCATTTTG | X1: 2359-20f | CAGGTCCCCTAGAAGAAGAA |
| CR1: 2502-21r | AATAAAGCCCAGTAAAGTTCC | SR: 861-23r | ACCCATCTTTTTGTTTTGTAG | X2: 2934-19r | CTGGTGATCGGGAAAGAAT |
| CF2: 1746-22r | GAGGAGATTAGGTTAAAGGTCT | X3: 725-18f | CAGTTATATGGATGATGT | X4: 1797-19r | CCAATTTATGCCTACAGCC |
| | | X5: 867-22f | CATTTGTTCAGTGGTTCGTAGG | X6: 1407-21r | CAGGATCCAGTTGGCAGCACA |
| Additional primer used for sequencing | Fragment B | SS: 3197-21f | CTCAGGCCATGCAGTGGAACT |

Results and Discussion

A recombinant genotype is dominant in Tibet

Fragments A and B of 26 randomly chosen extracted DNAs were amplified and sequenced. The sequence difference within fragment A was from 0 to 2.53% (pair-wise), and that within fragment B was from 0 to 1.67%, except for one isolate (Tibet705), which had a divergence of 5.57 to 6.90% from the others. Since most HBV sequences from different regions were highly homologous to each other, a dominant HBV genotype in Tibet was thus hypothesized. The isolate Tibet705 was classified as a genotype different from the dominant one.

Using these 26 sequences and the corresponding regions from 23 complete HBV nucleotide sequences from GenBank, we obtained two phylogenetic trees based on the surface antigen gene from fragment B and the core gene from fragment A (Fig. 1a, b). The HBV sequences clustered with genotype D in the trees based on the surface antigen gene, except for the isolate Tibet705, which clustered with genotype C. However, in trees based on the core gene, all sequences clustered with genotype C. The phylogenetic trees thus revealed an unknown recombinant HBV strain, which is prevalent in Tibet.

The complete genomes of Tibet127 and Tibet705 (accession nos AY057948 and AY057947), representative of the two groups, were sequenced. Both Tibet127 and Tibet705 were 3215 bp in length. Phylogenetic analysis classified both complete viral genomes into genotype C, with bootstrap values of 100% (Fig. 1c), although the surface antigen gene of
HBV recombinant in Tibet

Tibet127 was more similar to genotype D (divergence 1.5%) than genotype C (divergence 5-2%). The sequence divergence of the complete genomes and each HBV ORF was compared for Tibet127 and Tibet705 against the previously reported HBV genotypes (Table 2).

SimPlot and bootscanning analyses were applied to determine the possible recombination in Tibet127. It was suggested that in the Tibet127 isolate, the pre-S2/S gene and part of the P gene were from genotype D, whereas the rest of the viral genome was from genotype C (Fig. 2). The recombination spots were approximately at nt 50 (5' of pre-S2) and nt 1450 (3' end of P gene), as shown in Fig. 2. In agreement with this, the phylogenetic tree based on nt 51–1450 classified Tibet127 into genotype D with a bootstrap value of 100% (Fig. 3a), while in the tree based on nt 1451–50, the isolate was clustered with genotype C (Fig. 3b).

Accumulated evidence has suggested that recombination between viruses might be a relatively frequent event (Georgi-Geisberger et al., 1992; Bollyky et al., 1996; Bowyer & Sim, 2000; Hannoun et al., 2000; Morozov et al., 2000; Sugauchi et al., 2001). Although the mechanism is unknown, findings have supported the existence of a non-random mechanism. Three recombination hot spots in the vicinity of DR1 (direct repeat 1; nt 1800), the 3' end of the core region (nt 2359) and within the 3' end of the S gene were identified by analysing mosaic sequences in HBV (Bowyer & Sim, 2000; Morozov et al., 2000), with genotype D containing mosaics of genotype A and genotype B containing mosaics of genotype C. In the other reports, an aberrant genotype from Vietnam revealed a recombinant HBV strain (Hannoun et al., 2000), which showed recombination between genotypes A and C. Furthermore, strains from Australian aborigines showed that the complete HBV genome was similar to genotype C (6-9%), while the S gene differed greatly with the same genotype. The C/D hybrid genotype in this study differs from those reported recombinants. Moreover, the recombination sites of this C/D hybrid were at nt 50 and nt 1450, thus also differing from the

Table 2. Nucleotide differences (%) between the Tibet127 and Tibet705 isolates and sequences representing genotypes A–F

<table>
<thead>
<tr>
<th>Comparison of Tibet127 with genotype:</th>
<th>Comparison of Tibet705 with genotype:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Complete genome sequence</td>
<td>8.8</td>
</tr>
<tr>
<td>Small S</td>
<td>4.0</td>
</tr>
<tr>
<td>S ORF</td>
<td>6.5</td>
</tr>
<tr>
<td>C ORF</td>
<td>9.9</td>
</tr>
<tr>
<td>X ORF</td>
<td>5.2</td>
</tr>
<tr>
<td>P ORF</td>
<td>8.3</td>
</tr>
</tbody>
</table>
C. Chaoyin and others

Fig. 2. Nucleotide similarity comparison of isolate Tibet127 with consensus sequences representing each of the genotypes A–F, using SimPlot bootstrap analysis. The figure indicates that the isolate Tibet127 has high similarity with genotype D from nt 50–1450, while the isolate has high similarity with genotype C from nt 1450–50. Nt 50 and nt 1450 are therefore the recombination sites of this C/D hybrid.

Fig. 3. Phylogenetic analysis of each recombinant region of Tibet127, using the same methods as outlined in the legend to Fig. 1. (a) Nt 51–1450; (b) nt 1451–50.

Fig. 4. HBsAg sequences of serotypes ayw2, adw and adrq+ from residue 101 to 180.

Serotype analysis of Tibetan HBV strains

The amino acid residues specifying $d/y$ and $w/r$ were at positions 122 and 160 of the HBsAg (Okamoto et al., 1988). By comparing the amino acid sequences covering residues 101–180 of the HBsAg, 25 HBV strains showed the ayw2 serotype except for strain Tibet705, which was in the adw serotype based on Lys$^{122}$ and Lys$^{160}$ and differed from adrq+ only in residue 160 caused by a G → A transition at nt 633 (Fig. 4).
The HBV serotypes were consistent with a previous report that the serotype spread in Tibet was ayw (Luo et al., 1993).

Most ayw serotypes are grouped in genotypes B and D. However, our study indicated that the ayw strains isolated from Tibetans were grouped in genotype C, probably due to the recombination between the genotypes C and D.

Conclusion

We report for the first time the complete genome sequences of HBV strains isolated from the HBsAg-positive serum of Tibetans, which reveal that the dominant HBV genotype in Tibet is a C/D recombinant virus. These results may provide useful information to studies of the phylogenetic origin of the virus recombination, the contribution of the virus genotype to vaccine effects and clinical significance, and therefore the causes of the high HBV infection rate in the highland areas of Tibet.

The project was supported by a grant (No. 97-69) from the Science and Technology Bureau of Tibet, China, and a grant (970231006) from the National Science Foundation of China. Cui Chaoyin and Shi Jinxiu contributed equally to this work.

References


Received 9 February 2002; Accepted 22 March 2002