

A new intertype recombinant between genotypes C and D of hepatitis B virus identified in China

Zhanhui Wang,^{1,2} Zhihua Liu,² Guobing Zeng,² Shujuan Wen,² Yipeng Qi,¹ Shiwu Ma,² Nikolai V. Naoumov³ and Jinlin Hou^{2,3}

¹College of Life Science, Wuhan University, Wuhan, China

²Department of Infectious Diseases, Nanfang Hospital, Guangzhou 510515, China

³Institute of Hepatology, University College London, London WC1E 6HX, UK

Correspondence

Jinlin Hou

jihou@fimmu.com

Zhanhui Wang

wangzh@fimmu.com

Received 19 November 2004

Accepted 7 January 2005

Hepatitis B virus (HBV) genotypes have a characteristic geographical distribution. More than 90% of chronic HBV patients in China are infected with genotypes B or C. Here, eight HBV isolates that were initially classified as genotype D by PCR-restriction fragment length polymorphism analysis were analysed in detail. The complete HBV genome was sequenced and compared with 32 sequences retrieved from GenBank, representing HBV genotypes A–G. Phylogenetic analysis of the S gene (nt 10–800) classified all eight isolates as genotype D. However, phylogenetic analyses of nt 800–10 and the open reading frames (ORFs) of the precore/core and X genes classified all eight isolates as genotype C. This discordance between phylogenetic trees reconstructed on different ORFs suggested that intertype recombination has occurred in all eight isolates. By using the SIMPLOT program, the site of recombination with genotype D was located in the preS2/S region, spanning nt 10–799 in seven of eight isolates and nt 10–1499 in the other isolate. These results demonstrate that intertype recombination should be considered as a type of variation that increases the genetic diversity of HBV. Hybrids of different HBV genotypes might exhibit specific virological properties and their significance in the diagnosis and management of chronic hepatitis B deserves further investigation.

INTRODUCTION

Hepatitis B virus (HBV) has a worldwide distribution, with an estimated 350 million individuals having chronic HBV infection (Lee, 1997). This is of particular concern in China, the largest country in Asia, where the prevalence of chronic HBV infection is between 8 and 20% of the general population, with an estimated 130 million hepatitis B surface antigen (HBsAg) carriers (Luo *et al.*, 1996). The genome of HBV is a partially double-stranded circular DNA molecule of approximately 3200 bp that encodes four overlapping open reading frames (ORFs). Based on nucleotide sequence divergence over the whole genome of greater than 8%, HBV has been classified into eight genotypes, A–H (Okamoto *et al.*, 1988; Norder *et al.*, 1994; Stuyver *et al.*, 2000; Arauz-Ruiz *et al.*, 2002). Genotype A has two subgenotypes (Bowyer *et al.*, 1997; Kramvis *et al.*, 2002; Sugauchi *et al.*, 2004), renamed recently as A1 and A2 (Kimbi *et al.*, 2004). Genotype B has also been classified into two subgenotypes: Ba (for Asia) and Bj (for Japan) (Sugauchi *et al.*, 2002).

One of the most characteristic features of the eight currently known HBV genotypes is their distinct geographical distribution. Genotype A is mainly prevalent in north-western Europe and North America (Miyakawa & Mizokami, 2003). Genotypes B and C are highly prevalent in Asia. Genotype D has been found worldwide, but is predominant in the Mediterranean region. Genotype E is restricted almost entirely to West Africa and genotype F is found in Central and South America. Genotype G was found in Europe and the United States (Stuyver *et al.*, 2000).

In Asia, the most common HBV genotypes are B and C. Recently, we conducted a nationwide investigation, involving patients from nine provinces, to determine the distribution and virological characteristics of HBV genotypes in China (Zeng *et al.*, 2005). Four major genotypes, A, B, C and D, were found in this study and the prevalences of these four HBV genotypes were 1.2, 41, 52.5 and 4.3%, respectively (the remaining 1% of patients were infected with mixed genotypes of HBV). In this study, some genotype D isolates were found to have characteristics that were distinct from those of other genotype D isolates available in GenBank. Initially, these isolates were defined as genotype D by PCR-restriction fragment length polymorphism (RFLP) analysis, using an amplicon of the HBV surface (S)

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AY817509–AY817515 and AY800249.

gene. However, subsequent phylogenetic analysis based on entire nucleotide sequences revealed that they cluster with genotype C. Here, we demonstrate that genotype D isolates can be divided into two groups according to the region of recombination: one group possessed a recombination fragment of genotype D from nt 10 to 799, and the other had a longer recombination fragment of genotype D from nt 10 to 1499.

METHODS

Clinical material. Serum samples were collected from eight HBV DNA-positive patients who were resident in north-west China. All patients were diagnosed as chronic HBV carriers (serum HBsAg-positive for at least 6 months) and were seronegative for hepatitis C and hepatitis δ viruses. The serum samples were stored at -30°C until analysis.

HBV DNA preparation and amplification. Total DNA was extracted from 50 μl serum by using a DNA extractor kit (Huayin Inc.). DNA pellets were resuspended in 20 μl distilled water and 5 μl was used as a template for HBV DNA amplification. The PCR was performed in a 96-well cyler (GeneAmp PCR System 9700; Applied Biosystems) and in a 50 μl reaction volume containing 2 U LA *Taq* (TaKaRa). The PCR primers were P1 and P2 as described by Günther *et al.* (1995) after modification of the restriction enzyme-cleavage site from *HindIII* to *Sall*. This method has been shown to amplify the full-length HBV genome. The cycling conditions were initial denaturation at 94°C for 1 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30 s and extension at 72°C for 3.5 min. Amplicons (6 μl) were analysed by electrophoresis on 1.5% agarose gel, stained with ethidium bromide and observed under UV light.

Cloning and sequencing of the full-length genome. The PCR products were purified with a QIAquick gel extraction kit (Qiagen) and cloned into vector pMD 18-T (TaKaRa) by using standard cloning techniques. White colonies were picked and grown in Luria-Bertani medium with ampicillin ($100\ \mu\text{g ml}^{-1}$). The correct insert size was confirmed by using PCR and the restriction enzyme *Sall* (Promega). DNA sequencing analysis of the correct recombinants was performed with an ABI 3730 automated DNA sequencer (Applied Biosystems). For each patient, five recombinants were sequenced over the S and core (C) genes to confirm the genotype and one recombinant was selected for sequencing of the complete genome.

Phylogenetic analysis. All HBV DNA sequences were aligned by using the CLUSTAL W program (version 1.7; EMBL) and the alignment was confirmed by visual inspection. The alignments were fed into the PHYLIP software package, version 3.5c (Felsenstein, 1993). Genetic distances were estimated by the Kimura two-parameter matrix and phylogenetic trees were constructed by the neighbour-joining method; the reliability of topologies was estimated by performing bootstrap resampling and reconstruction with 1000 replicates, then the CONSENSE program in the PHYLIP package was used to compute a consensus tree.

To detect sequences with conflicting phylogenetic positions, phylogenetic trees were reconstructed based on entire genome, the S region from nt 10 to 800, the remaining fragment from nt 800 to 10 and the preC/C and X genes.

Recombination investigation. Recombination was searched for with the SIMPLOT program (available at <http://sray.med.som.jhmi.edu/RaySoft/SimPlot>) and bootscanning analysis (Robertson *et al.*,

1995; Lole *et al.*, 1999; Sugauchi *et al.*, 2002). The SIMPLOT program, version 2.5, was used to identify phylogenetically informative sites supporting alternative tree topologies. This was performed by considering four sequences at a time: one putative recombinant sequence, two reference sequences of the original C and D genotypes and one sequence of genotype F as a known outgroup. Each informative site supports one of three possible phylogenetic relationships among the four taxa. Bootscanning and cluster analysis maximizing χ^2 were used to identify the breakpoints in the intergenotypic recombinants. *P* values for the subsequent division of the sequence into genotypes were calculated by using Fisher's exact test.

RESULTS

Nucleotide sequences and phylogenetic analysis

Based on PCR-RFLP using an amplicon of the HBV S gene (Zeng *et al.*, 2004), the HBV DNA isolates from all eight patients were defined as genotype D (data not shown). However, DNA sequencing of the S and C genes in all five clones from each patient showed genotype C for the C gene and genotype D for the S gene. The complete nucleotide sequences of the eight isolates were then determined and every isolate was found to possess a full genome length of 3215 bp. We compared the full HBV DNA nucleotide sequences of these eight isolates with 32 HBV DNA isolates retrieved from GenBank, representing HBV genotypes A–G. All eight HBV/D isolates in the present study clustered on a branch within genotype C (Fig. 1a). Complete nucleotide sequence AY817511 was very similar to Tibet127 (accession no. AY057948), an HBV C/D hybrid subtype found in Tibet (Cui *et al.*, 2002). Phylogenetic analysis based on the S gene (nt 10–800) classified all eight isolates into genotype D (Fig. 1b). However, in phylogenetic analyses based on HBV DNA sequences between nt 800 and 10, as well as the preC/C and X genes (data not shown), all eight isolates clustered with genotype C (Fig. 1c), a result similar to those of the phylogenetic analysis of the entire HBV genome. Comparison of the eight new isolates with the full genomes of genotypes A–F showed no more than 5.4% divergence from genotype C. Comparison over the C and X genes only showed even less divergence from genotype C: not more than 3%. Instead, over the S gene, similarity to genotype D was 97.9–98.4%, which was higher than that to genotype C (Table 1). This analysis revealed the presence of a potential new subgenotype of HBV genotype C.

Identification of the putative recombination site

The clustering of sequences in discordant positions according to phylogenetic analyses was an indication of recombination. Next, SIMPLOT and bootscanning analyses were applied to determine the possible recombination sites in these eight genotype C isolates. Fig. 2(b) shows the genetic distances between AY817512 and the other two reference sequences, HBV/C (GenBank accession no. AB033550) and HBV/D (accession no. AF280817), over the complete genome, with a window size of 200 bp and a step size of 50 bp. The distance plots of the complete sequence showed regions of higher similarity to genotype C, alternating with

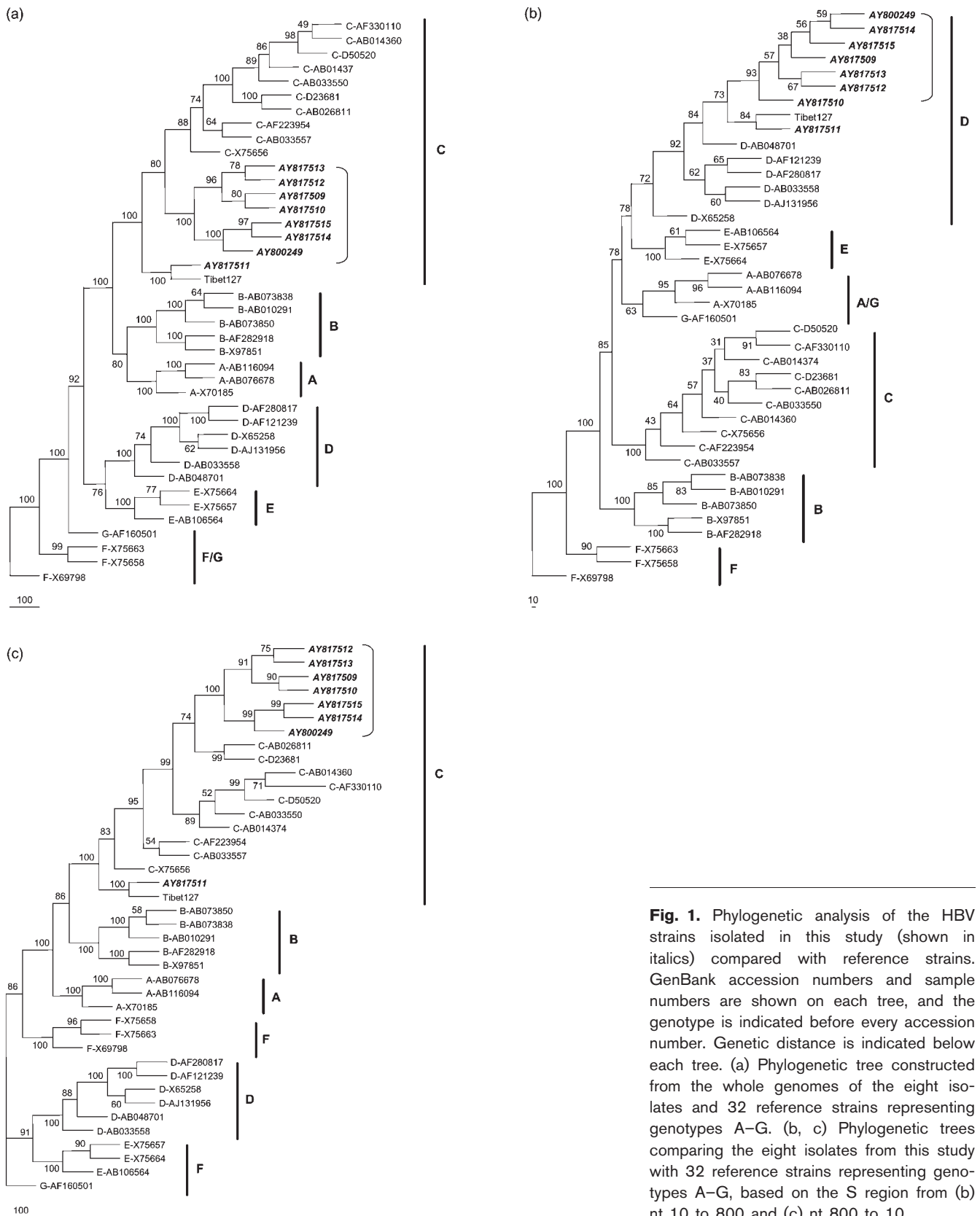


Fig. 1. Phylogenetic analysis of the HBV strains isolated in this study (shown in italics) compared with reference strains. GenBank accession numbers and sample numbers are shown on each tree, and the genotype is indicated before every accession number. Genetic distance is indicated below each tree. (a) Phylogenetic tree constructed from the whole genomes of the eight isolates and 32 reference strains representing genotypes A–G. (b, c) Phylogenetic trees comparing the eight isolates from this study with 32 reference strains representing genotypes A–G, based on the S region from (b) nt 10 to 800 and (c) nt 800 to 10.

Table 1. Comparison of divergence (%) in nucleotide sequences between the eight new isolates and reference sequences representing genotypes A–F

Each of the eight new isolates was compared with reference sequences representing genotypes A–F within the complete genome and the S, C and X genes. GenBank accession numbers of the representative isolates are AB116094, AB073838, D50520, AF280817, X75657 and X69798 for genotypes A–F, respectively.

Sequence	A	B	C	D	E	F
Complete genome	9.3–8.5	9.8–9.1	5.4–3.4	12.8–11.2	9.9–9.6	13.3–12.8
S gene	7.0–6.1	7.6–7.2	6.5–6.2	2.1–1.6	5.5–4.9	9.6–8.9
C gene	9.8–9.1	10.2–9.2	2.3–1.6	8.8–8.0	9.7–9.1	10.5–9.9
X gene	7.7–6.2	6.2–4.5	2.8–1.3	6.0–4.7	5.8–4.5	11.0–8.8

the regions of higher similarity to genotype D within the S gene and overlapping the P gene. In order to locate the site of possible recombination, bootscanning analysis was carried out by using the reference sequences of genotypes C and D for comparison and genotype F as the outgroup. Fig. 2(a) depicts the results of the bootscanning analysis of AY817512, which revealed that the S gene of AY817512 was substituted by the corresponding part of genotype D. The locations of the recombination breakpoints were estimated at nt 10 and 799.

The recombination region of AY817511 was different from that of the other seven isolates. Fig. 3 shows that, for AY817511, the change in genotype occurred within the large part of the P ORF, overlapping the S ORF and the amino terminal of the X ORF, and the breakpoints were located at nt 10 and 1499.

When the nucleotide sequences of the eight new isolates, along with the representative sequences of genotypes C and D, were aligned, it was clear that the change of sequence characteristics in the eight isolates occurred at nt 10 and 799 and at nt 10 and 1499 for AY817511. Comparison of the amino acid sequences between residues 120 and 170 of HBsAg showed that all eight isolates belong to the *ayw2* serotype, which is often grouped as genotype D.

DISCUSSION

Previous reports have described HBV strains of certain genotypes in which part of the viral genome is replaced by the corresponding part from another HBV genotype, as a result of recombination (Bollyky *et al.*, 1996; Bowyer & Sim, 2000). More recently, accumulating data have suggested that DNA recombination is a significant and relatively frequent event in the evolution of HBV. The existence of hybrids between HBV genotypes B and C (B/C) was found frequently in South-East Asia, and the recombination region was located in the preC/C gene (Morozov *et al.*, 2000; Sugauchi *et al.*, 2002, 2003; Luo *et al.*, 2004). Recombination between genotypes A and D was described in HBV strains in Italy and South Africa (Morozov *et al.*, 2000; Owiredu *et al.*, 2001). Analysis of an aberrant HBV genotype from Vietnam revealed a recombination between

genotypes A and C (Hannoun *et al.*, 2000). Cui *et al.* (2002) reported a hybrid of HBV genotypes C and D that was isolated in Tibet and identified the recombination sites at nt 50 and 1450. In the present study, we found a new C/D hybrid that differs from the previously reported recombinants. By SIMPLOT and bootscanning analysis, the breakpoints of the new C/D hybrid were located at nt 10 and 799, which are also different from the breakpoints of the C/D hybrid identified in Tibet. All of the eight new isolates were C/D hybrids with a distinct recombinant region: seven of eight isolates had recombination with genotype D over the preS2/S region (nt 10–799), whilst one isolate (GenBank accession no. AY817511) was different, having recombination over a large part of the P gene overlapping with the S and X genes, identical to another C/D hybrid that was identified in Tibet. So far, HBV C/D hybrids have only been found in China. However, the authentic genotype D has a worldwide distribution – in southern Europe, the Americas and Australia. Yan *et al.* (2001) reported the complete genome of a genotype D isolate from Ningxia, China, and this isolate (accession no. AF280817) is the only reported authentic genotype D isolate found in China, which suggested that the prevalence of the authentic genotype D is very low in China. Further studies are needed to investigate the geographical distribution of C/D hybrids and to clarify whether these HBV recombinants are restricted to China.

Traditionally, the definition of HBV genotypes has been based on one of the following criteria: an intergroup divergence of 8% or greater over the complete genome sequence, or 4.1% or greater divergence of the surface-antigen gene (Okamoto *et al.*, 1988; Norder *et al.*, 1994). When these genotyping criteria were applied to determine the genotype of the eight isolates in the present study, we obtained conflicting results. Based on PCR-RFLP using an amplicon of the HBV S gene, these eight isolates were defined as genotype D; the same results were obtained with phylogenetic analysis of the S gene (nt 10–800). However, in phylogenetic analyses of nt 800–10 and the other two ORFs, preC/C and X, all eight isolates clustered with genotype C, as with the phylogenetic analysis of the entire genome. These results demonstrate that HBV isolates might not be genotyped correctly based only on the S gene, because of possible recombination events that may occur between

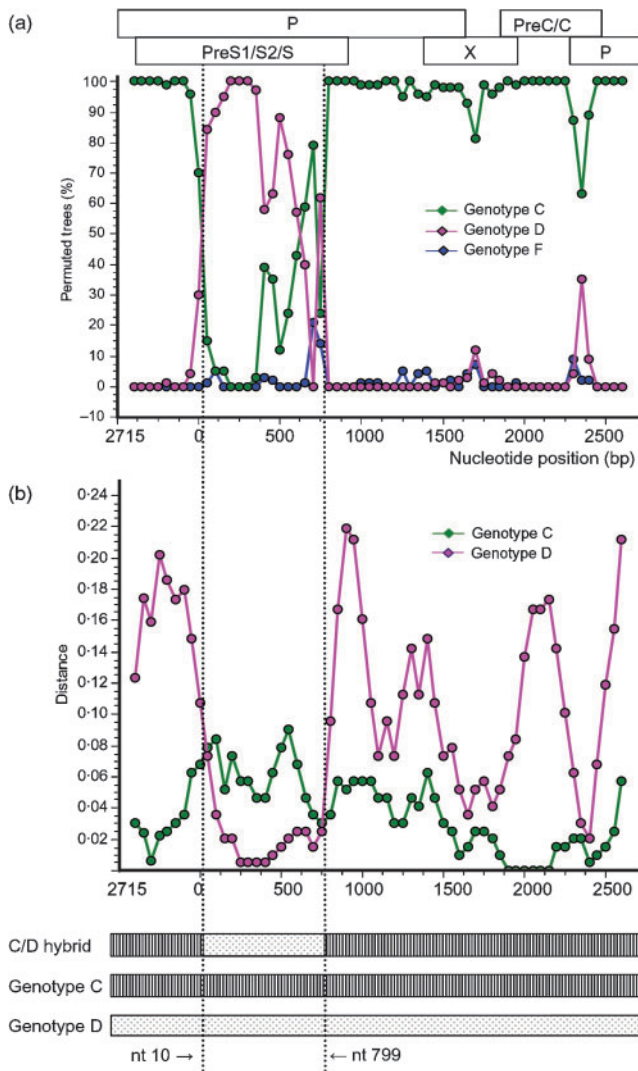


Fig. 2. (a) The location of the recombination event in isolate AY817512 was determined by using the SIMPLOT program and bootscanning analysis. Isolate AY817512 was compared with three representative isolates of HBV/C (GenBank accession no. AB033550), HBV/D (accession no. AF280817) and outgroup HBV/F (accession no. X75658) on the full genome with 200 bp window size, 50 bp step size, 100 bootstrap replicates, gapstrip on and neighbour-joining analysis. (b) Genetic distances compared among HBV/C (GenBank accession no. AB033550) and HBV/D (accession no. AF280817) with AY817512 over the complete genome, with a window size of 200 bp and a step size of 50 bp. The dotted vertical lines show the breakpoints of recombination.

different genotypes. Therefore, it is a challenge to the traditional genotyping criteria that an increasing number of HBV isolates have been identified as carrying mosaic genomes as a result of recombination with another genotype.

Intertypic recombinations of HBV strains have been

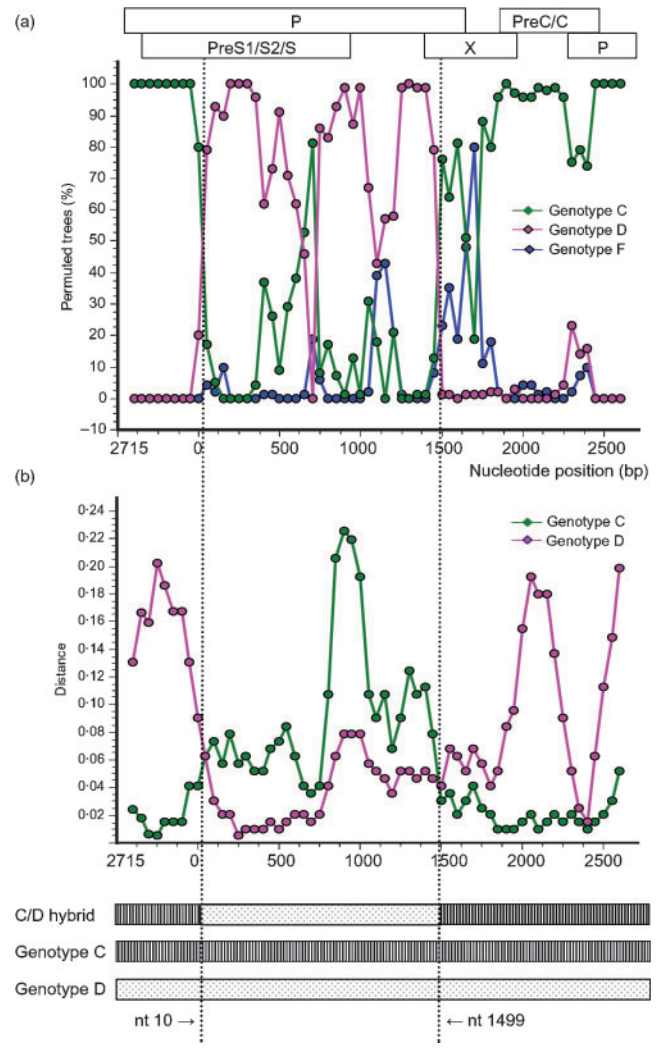


Fig. 3. (a) The location of the recombination event in isolate AY817511 was determined by using the SIMPLOT program and bootscanning analysis. Isolate AY817511 was compared with three representative isolates of HBV/C (GenBank accession no. AB033550), HBV/D (accession no. AF280817) and outgroup HBV/F (accession no. X75658) on the full genome with 200 bp window size, 50 bp step size, 100 bootstrap replicates, gapstrip on and neighbour-joining analysis. (b) Genetic distances compared among HBV/C (accession no. AB033550) and HBV/D (GenBank accession no. AF280817) with AY817511 over the complete genome, with a window size of 200 bp and a step size of 50 bp. The dotted vertical lines show the breakpoints of recombination.

described between different genotypes: B/C, A/D, C/D and A/C, etc. Comparison of these data indicates that some fragments of the HBV genome are prone to be replaced by the corresponding parts of an alternative genotype, e.g. the preC/C region in B/C hybrids and the preS2/S region in C/D hybrids. One possible mechanism may be immunological selection pressure, if the intertypic recombinant could facilitate evasion of immune surveillance.

Whether HBV genotypes differ in their immunogenicity and their immunodominant T-cell epitopes is currently unknown. The possibility for T-cell cross-reactivity between HBV genotypes also requires further investigation.

In this study, we identified a new C/D hybrid in north-west China. The new C/D hybrid was classified into genotypes D and C, based on the preS2/S region and complete genome, respectively. The geographical distribution and clinical significance of these new C/D hybrids, as well as their impact on HBV-specific T-cell reactivity, will be studied in the future.

ACKNOWLEDGEMENTS

This work was supported by a grant for the National Natural Science Foundation of China and Guangdong and the National Science Fund for Distinguished Young Scholars. The authors thank the doctors from the Provincial People's Hospital of Gansu, Lanzhou, Gansu, China, for collecting serum samples for the study.

REFERENCES

- Arauz-Ruiz, P., Norder, H., Robertson, B. H. & Magnius, L. O. (2002). Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *J Gen Virol* **83**, 2059–2073.
- Bollyky, P. L., Rambaut, A., Harvey, P. H. & Holmes, E. C. (1996). Recombination between sequences of hepatitis B virus from different genotypes. *J Mol Evol* **42**, 97–102.
- Bowyer, S. M. & Sim, J. G. M. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J Gen Virol* **81**, 379–392.
- Bowyer, S. M., van Staden, L., Kew, M. C. & Sim, J. G. M. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *J Gen Virol* **78**, 1719–1729.
- Cui, C., Shi, J., Hui, L., Xi, H., Zhuoma, Quni, Tsedan & Hu, G. (2002). The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol* **83**, 2773–2777.
- Felsenstein, J. (1993). PHYLIP (phylogeny inference package), version 3.5c. Department of Genetics, University of Washington, Seattle, USA.
- Günther, S., Li, B. C., Miska, S., Krüger, D. H., Meisel, H. & Will, H. (1995). A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J Virol* **69**, 5437–5444.
- Hannoun, C., Norder, H. & Lindh, M. (2000). An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *J Gen Virol* **81**, 2267–2272.
- Kimbi, G. C., Kramvis, A. & Kew, M. C. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J Gen Virol* **85**, 1211–1220.
- Kramvis, A., Weitzmann, L., Owiredu, W. K. B. A. & Kew, M. C. (2002). Analysis of the complete genome of subgroup A' hepatitis B virus isolates from South Africa. *J Gen Virol* **83**, 835–839.
- Lee, W. M. (1997). Hepatitis B virus infection. *N Engl J Med* **337**, 1733–1745.
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.
- Luo, K., Hou, J. & Liang, C. (1996). Hepatitis B virus infection and its genetic variations in China. *Chin J Infect Dis* **14**, 101–104.
- Luo, K., Liu, Z., He, H., Peng, J., Liang, W., Dai, W. & Hou, J. (2004). The putative recombination of hepatitis B virus genotype B with pre-C/C region of genotype C. *Virus Genes* **29**, 31–41.
- Miyakawa, Y. & Mizokami, M. (2003). Classifying hepatitis B virus genotypes. *Intervirology* **46**, 329–338.
- Morozov, V., Pisareva, M. & Groudinin, M. (2000). Homologous recombination between different genotypes of hepatitis B virus. *Gene* **260**, 55–65.
- Norder, H., Couroucé, A.-M. & Magnius, L. O. (1994). Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* **198**, 489–503.
- Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R. I., Imai, M., Miyakawa, Y. & Mayumi, M. (1988). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol* **69**, 2575–2583.
- Owiredo, W. K. B. A., Kramvis, A. & Kew, M. C. (2001). Molecular analysis of hepatitis B virus genomes isolated from black African patients with fulminant hepatitis B. *J Med Virol* **65**, 485–492.
- Robertson, D. L., Hahn, B. H. & Sharp, P. M. (1995). Recombination in AIDS viruses. *J Mol Evol* **40**, 249–259.
- Stuyver, L., De Gendt, S., Van Geyt, C., Zoulim, F., Fried, M., Schinazi, R. F. & Rossau, R. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol* **81**, 67–74.
- Sugauchi, F., Orito, E., Ichida, T. & 9 other authors (2002). Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J Virol* **76**, 5985–5992.
- 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.
- Sugauchi, F., Kumada, H., Acharya, S. A. & 10 other authors (2004). Epidemiological and sequence differences between two subtypes (Ae and Aa) of hepatitis B virus genotype A. *J Gen Virol* **85**, 811–820.
- Yan, L., Hou, J.-L., Guo, Y.-B., Chen, J.-J., Wang, Z.-H., Lin, Y.-L., Luo, K.-X. & Niu, Z.-Y. (2001). Determination and analysis of complete nucleotide sequence of Chinese genotype D hepatitis B virus: the first report. *Di Yi Jun Yi Da Xue Xue Bao* **21**, 848–851.
- Zeng, G.-B., Wen, S.-J., Wang, Z.-H., Yan, L., Sun, J. & Hou, J.-L. (2004). A novel hepatitis B virus genotyping system by using restriction fragment length polymorphism patterns of S gene amplicons. *World J Gastroenterol* **10**, 3132–3136.
- Zeng, G., Wang, Z., Wen, S. & 10 other authors (2005). Geographic distribution, virologic and clinical characteristics of hepatitis B virus genotypes in China. *J Viral Hepat* (in press).