Journal of General Virology (2002), **83**, 2773–2777. Printed in Great Britain

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] Tsedan[1] and Gengxi Hu[2]

[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,

Author for correspondence: Gengxi Hu (Fax 86 21 64718563. e-mail hgxgene@sunm.shcnc.ac.cn) and Tsedan (e-mail tsedan_2000@yahoo.com).

0001-8391 © 2002 SGM

Table 1. Primers used for HBV DNA amplification and sequencing

	Primers used for amplification	
Fragment A	CF1: 1719-21f	GTGTTTAAGGACTGGGAGGAG
	CR1: 2502-21r	AATAAAGCCCAGTAAAGTTCC
	CF2: 1746-22r	GAGGAGATTAGGTTAAAGGTCT
Fragment B	SF: 2798-21f	AACACATAGCGCCTCATTTTG
	SR: 861-23r	ACCCCATCTTTTTGTTTTGTTAG
Fragment I	X1: 2359-20f	CAGGTCCCCTAGAAGAAGAA
	X2: 2934-19r	CTGGTGATCGGGAAAGAAT
Fragment II	X3: 725-18f	CAGTTATATGGATGATGT
	X4: 1797-19r	CCAATTTATGCCTACAGCC
Fragment III	X5: 687-22f	CATTTGTTCAGTGGTTCGTAGG
	X6: 1407-21r	CAGGATCCAGTTGGCAGCACA
	Additional primer used for sequencing	
Fragment B	SS: 3197-21f	CTCAGGCCATGCAGTGGAACT

the main towns of the Tibet Autonomous Region. No abnormal liver function was observed in these students. All subjects included in the study were Tibetan. Twenty-six samples were randomly chosen for our study, with the average age of the subjects being 14·8 years old.

■ **HBV DNA preparation and amplification.** Serum samples were stored at −80 °C until analysis. Virus DNA was extracted from 50 µl serum treated with 1% NP-40 and 0·75% Tween 20. Two fragments corresponding to the HBV genome sequence nt 1746–2502, including the pre-C/C gene (fragment A), and nt 2798–861, including the pre-S1/pre-S2/S gene (fragment B), were amplified by PCR. Fragment A was amplified using primers CF1 and CR1, followed by a semi-nested reaction using primers CF2 and CR1. The complete genomes of two HBV isolates were amplified using the primers listed in Table 1. A typical amplification was performed in a 30 µl reaction volume containing 2 µl extracted DNA and *Taq* polymerase for 35 cycles at 94 °C for 1 min, 58 °C for 1 min and 72 °C for 1 min. Standard precautions to avoid contamination during PCR were taken, including a negative control serum included in each run.

■ **Sequence determination.** After purification on Wizard PCR Preps DNA purification resin (Promega), PCR products were bidirectionally sequenced with the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) using the PCR primers. The PCR products of fragment B were about 1200 bp, so an internal sequencing primer, SS, was designed in addition to the amplification primers SF and SR for sequencing (Table 1). Sequencing was performed on an automated DNA sequencer ABI 377 (PE Applied Biosystems).

■ **Data analysis.** The nucleotide sequences of the Tibetan HBV strains were compared with those of the 23 reference HBV strains obtained from GenBank, representing each of the six genotypes, A–F. Phylogenetic trees were constructed with the MEGA program version 2.1 (Kumar *et al*., 1994), using the Kimura two-parameter matrix and the neighbour-joining method. To confirm the reliability of the phylogenetic tree analysis, bootstrap resampling and reconstruction were carried out 500 times. Recombination was investigated by SimPlot (Lole *et al*., 1999) (distributed by the author, Stuart Ray, at http://www.welch.jhu.edu/~sray) and bootscanning (Salminen *et al*., 1995) analysis.

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

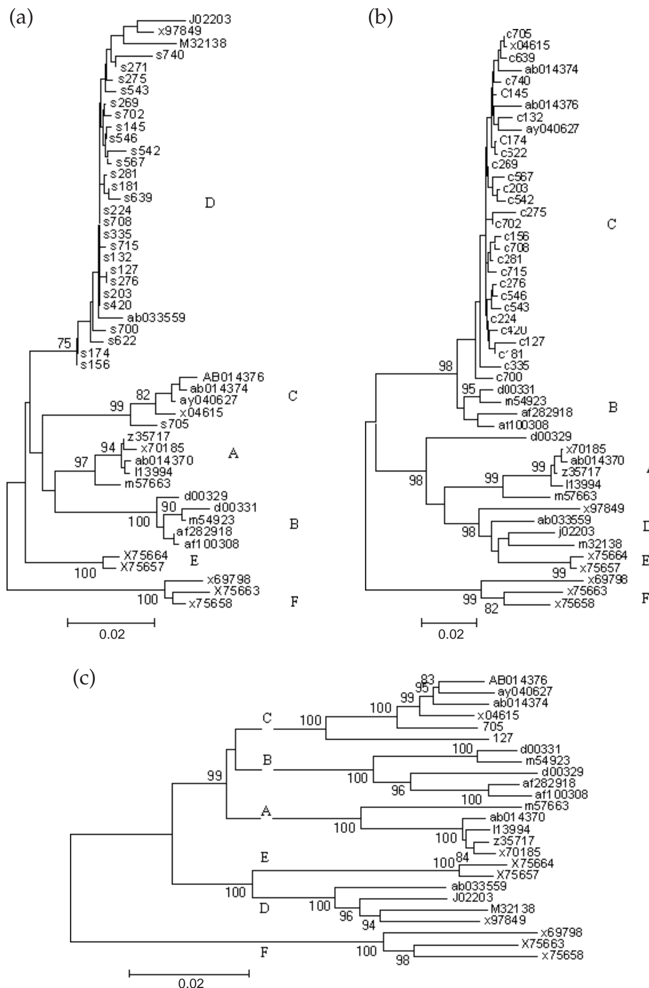

Fig. 1. Phylogenetic analysis of the HBV strains isolated from Tibet compared with 23 reference strains. Genetic distances were estimated by the Kimura two-parameter matrix and polygenetic trees were constructed by the neighbour-joining method. Accession numbers and sample numbers are indicated on the tree. Bootstrap values are shown along each main branch. The regions included in the analysis were: (a) the ORF of the surface antigen gene, not including pre-S1 and pre-S2; (b) the ORF of the core gene, not including the pre-core region; (c) the complete genome.

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 these 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

	Comparison of Tibet127 with genotype:						Comparison of Tibet705 with genotype:					
	A	B	C	D	E	F	A	B	C	D	E	F
Complete genome sequence	8·8	9·1	***5·4***	7·9	10·6	14·8	8·0	8·9	***2·9***	10·8	10·6	14·2
Small S	4·0	4·9	5·2	***1·5***	4·0	7·3	5·1	5·6	***1·6***	5·0	5·4	8·5
S ORF	6·5	9·0	***4·9***	6·5	8·8	14·0	5·3	9·1	***3·1***	9·6	9·6	14·9
C ORF	9·9	4·5	***1·9***	10·2	10·9	10·9	10·3	3·8	***1·0***	10·3	11·3	11·4
X ORF	5·2	5·7	***3·0***	4·0	4·9	10·2	6·3	8·0	***4·2***	7·2	7·0	11·5
P ORF	8·3	10·2	***6·3***	7·4	10·2	14·6	7·3	9·8	***3·1***	10·7	10·6	14·4

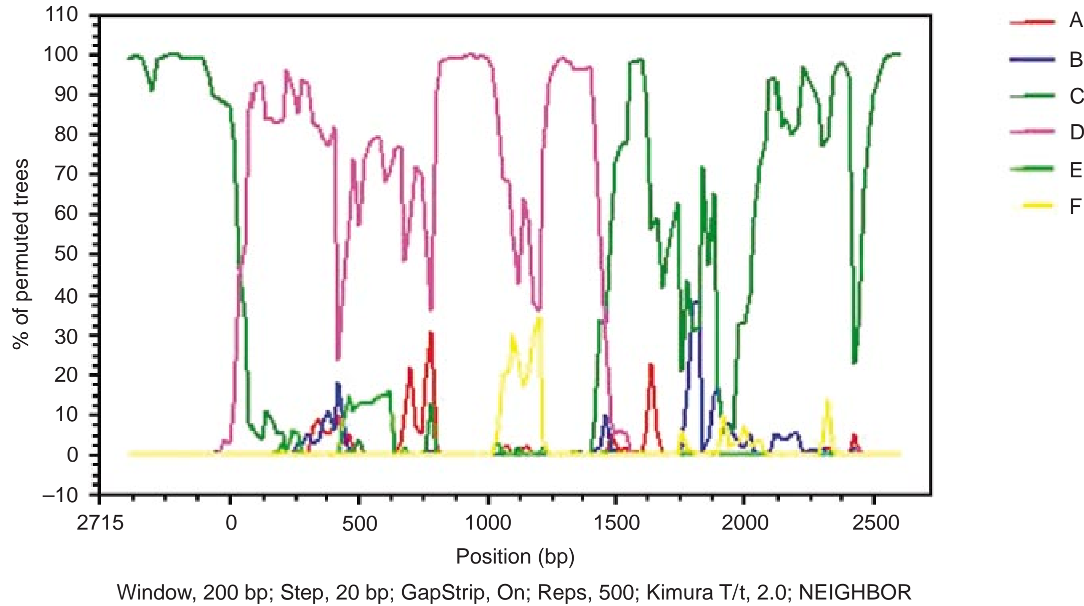

Fig. 2. Nucleotide similarity comparison of isolate Tibet127 with consensus sequences representing each of the genotypes A–F, using SimPlot bootstrap analysis. The figure indicates that the isolate Tibet127 has high similarity with genotype D from nt 50–1450, while the isolate has high similarity with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

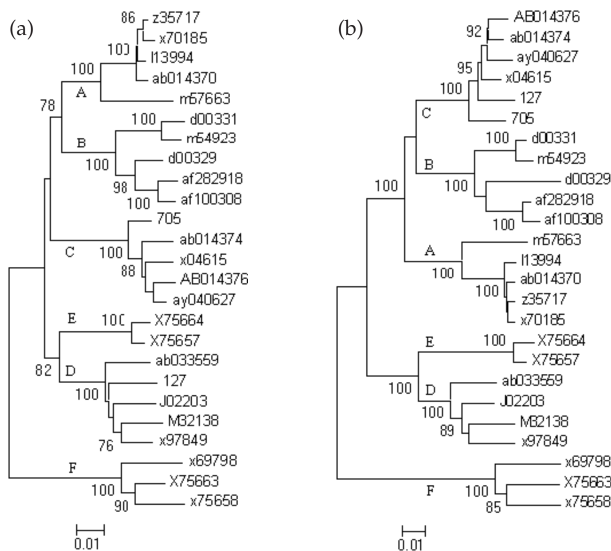

Fig. 3. Phylogenetic analysis of each recombinant region of Tibet127, using the same methods as outlined in the legend to Fig. 1. (a) Nt 51–1450; (b) nt 1451–50.

sites in the other recombinants, and this hybrid could thus be an offspring of an early recombination event. This might be an interesting characteristic of the HBV strains in Tibet, and suggests that the mechanism of recombination may be more complicated than anticipated.

It is interesting that 96 % of the HBV isolates analysed from Tibet were C/D recombinants. Genotype C is one of the genotypes that prevails in the Chinese Hans, whereas genotype D is predominant in the Mediterranean region. More studies are needed to determine whether this type of recombinant can adapt to the special environment of the highland areas and to the specific genetic background of the Tibetans. Analysis of additional sequences from Tibet and its neighbouring areas would be helpful to investigate the possible origin of this C/D recombinant and to learn how recombinant viruses differ in their pathogenicity.

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 *ayw*2 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).

```
          101       110          120          130          140          150          160          170          180
adw       QGMLPVCPLI|PGSTTTSTGP|CKTCTTPAQG|NSMFPSCCCT|KPTDGNCTCI|PIPSSWAFAK|YLWEWASVRF|SWLSLLVPFV|
ayw2      ---------- ---S------ -R-------- T--Y------ --S------- --------G- F------A-- ----------
Tibet127  ---------- ---S------ -R-------- T--Y------ --S------- --------G- F------A-- ----------
adrq+     ---------L --TS------ -----I---- T--------- --S------- ---------R F--------- ----------
Tibet705  ---------L --TS------ -----I---- T--------- --S------- ---------- F--------- ----------
```

Fig. 4. HBsAg sequences of serotypes *ayw2*, *adw* and $adrq^{+}$ from residue 101 to 180.

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

Bollyky, P. L., Rambaut, A., Harvey, P. H. & Holmes, E. C. (1996). Recombination between sequences of hepatitis B virus from difference genotypes. *Journal of Molecular Evolution* **42**, 97–102.

Bowyer, S. M. & Sim, J. G. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *Journal of General Virology* **81**, 379–392.

Georgi-Geisberger, P., Berns, H., Loncarevic, I. F., Yu, Z. Y., Tang, Z. Y., Zentgraf, H. & Schroder, C. H. (1992). Mutations on free and integrated hepatitis B virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. *Oncology* **49**, 386–395.

Hannoun, C., Norder, H. & Lindh, M. (2000). An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *Journal of General Virology* **81**, 2267–2272.

Kane, M. A. (1996). Globe status of hepatitis B immunization. *Lancet* **348**, 696.

Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Computer Applications in the Biosciences* **10**, 189–191.

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. *Journal of Virology* **73**, 152–160.

Luo, K., He, C. & He, H. (1993). The investigation of hepatitis B virus in Han, Tibetan, Dai, Yao, Uighul, Mongolia and Li populations. *Chinese Journal of Epidemiology* **14**, 266–270.

Morozov, V., Pisareva, M. & Groudinin, M. (2000). Homologous recombination between different genotypes of hepatitis B virus. *Gene* **260**, 55–65.

Norder, H., Hammas, B., Lofdahl, S., Courouce, A. M. & Magnius, L. O. (1992). Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *Journal of General Virology* **73**, 1201–1208.

Norder, H., Courouce, 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* **98**, 489–503.

Okamoto, H., Imai, M., Shimozaki, M., Hoshi, Y., Iizuka, H., Gotanda, T., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1986). Nucleotide sequence of a cloned hepatitis B virus genome, subtype ayr: comparison with genomes of the other three subtypes. *Journal of General Virology* **67**, 2305–2314.

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. *Journal of General Virology* **69**, 2575–2583.

Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombination of HIV type 1 by bootscanning. *AIDS Research and Human Retroviruses* **11**, 1423–1425.

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. *Journal of General Virology* **81**, 67–74.

Stuyver, L., Locarnini, S. A., Lok, A., Richman, D. D., Carman, W. F., Dienstag, J. L. & Schinazi, R. F. (2001). Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology* **33**, 751–757.

Sugauchi, F., Mizokami, M., Orito, E., Ohno, T., Kato, H., Suzuki, S., Kimura, Y., Ueda, R., Butterworth, L. A. & Cooksley, W. G. (2001). A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *Journal of General Virology* **82**, 883–892.

Zhu, B., Luo, K. X., Hu, Z. Q., Hou, J. L., Zhou, F. Y. & Feng, Y. R. (1999). Establishment of a method for classification of HBV genome and its application. *Chinese Journal of Experimental and Clinical Virology* **4**, 309–313.

Received 9 February 2002; Accepted 22 March 2002