Journal of General Virology (2002), **83**, 2059–2073. Printed in Great Britain

Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America

Patricia Arauz-Ruiz,[1,2] Helene Norder,[1] Betty H. Robertson[3] and Lars O. Magnius[1]

[1] Department of Virology, Swedish Institute for Infectious Disease Control, SE-171 82 Stockholm, Sweden

[2] Louisiana State University-International Centre for Medical Research and Training, San José, Costa Rica

[3] Laboratory Branch, Division of Viral Hepatitis, National Center for Infectious Diseases, Center for Disease Control and Prevention, Atlanta, GA 30333, USA

The complete genomes were sequenced for ten hepatitis B virus (HBV) strains. Two of them, from Spain and Sweden, were most similar to genotype D, although encoding *d* specificity. Five of them were from Central America and belonged to genotype F. Two strains from Nicaragua and one from Los Angeles, USA, showed divergences of 3·1–4·1% within the small S gene from genotype F strains and were recognized previously as a divergent clade within genotype F. The complete genomes of the two genotype D strains were found to differ from published genotype D strains by 2·8–4·6%. Their S genes encoded Lys^{122}, Thr^{127} and Lys^{160}, corresponding to the putative new subtype *adw3* within this genotype, previously known to specify *ayw2*, *ayw3* or, rarely, *ayw4*. The complete genomes of the three divergent strains diverged by 0·8–2·5% from each other, 7·2–10·2% from genotype F strains and 13·2–15·7% from other HBV strains. Since pairwise comparisons of 82 complete HBV genomes of intratypic and intertypic divergences ranged from 0·1 to 7·4% and 6·8 to 17·1%, respectively, the three sequenced strains should represent a new HBV genotype, for which the designation H is proposed. In the polymerase region, the three strains had 16 unique conserved amino acid residues not present in genotype F strains. So far, genotype H has been encountered in Nicaragua, Mexico and California. Phylogenetic analysis of the complete genomes and subgenomes of the three strains showed them clustering with genotype F but forming a separate branch supported by 100% bootstrap. Being most similar to genotype F, known to be an Amerindian genotype, genotype H has most likely split off from genotype F within the New World.

Introduction

Hepatitis B virus (HBV) is the prototype member of the family *Hepadnaviridae*. It has a compact, circular DNA genome of about 3·2 kb in length with four overlapping open reading frames (ORFs). The overlapping ORFs impose constraints on possible nucleotide substitutions and to variable substitution rates for different genomic regions (Yang *et al.*, 1995). The rate at which HBV sequences mutate is, however, uncertain; one study in an HBV e antigen (HBeAg)-positive carrier suggests substitution rates in the order of 10^{-5} per year per site (Okamoto *et al.*, 1987a).

HBV strains are classified into seven main genomic groups or genotypes, designated A–G, and arbitrarily defined by an intergroup divergence of more than 8% based on complete genomes (Norder *et al.*, 1992b; Okamoto *et al.*, 1988). Genotype A is prevalent in Northern and Central Europe but is also common in North America and sub-Saharan Africa. Genotypes B and C are confined to Asia. Genotype D is widespread but is the predominant genotype in the Mediterranean region, while genotype E is found mainly in West Africa. Genotype F shows the highest divergence among the genotypes and is indigenous to aboriginal populations of the Americas (Norder *et al.*, 1993). The newly described genotype G has been found in the USA and France (Stuyver *et al.*, 2000). Furthermore, some genotypes have been split into subgroups

Author for correspondence: Lars Magnius.
Fax +46 8 33 72 72. e-mail lars.magnius@smi.ki.se

The sequences reported in this paper will appear in the DDBJ/EMBL/GenBank databases under the accession numbers AY090452–61.

Table 1. HBV strains retrieved from GenBank used in this study

Strains cited by Norder *et al.* (1996) and Arauz-Ruiz *et al.* (1997a) are not included in this table.

Strain	Genotype	Accession no.	Reference
A317	A	Z72478	Schories *et al.* (2000)
JL1	B	AJ131133	
ASA-FH1	B	D50521	Asahina *et al.* (1996)
Pat4-*adw*	B	X97850	Alexopoulou *et al.* (1996)
Pat6-*adw*	B	X97851	Alexopoulou *et al.* (1996)
PF1	B	X98077	Pult *et al.* (1997)
pSK619	C	AB033553	Okamoto *et al.* (1987b)
HBAustRC	C	AB048704	Sugauchi *et al.* (2001)
HBAustSJ	C	AB048705	Sugauchi *et al.* (2001)
11141-rec	C	AF241407	Hannoun *et al.* (2000)
6871-rec	C	AF241409	Hannoun *et al.* (2000)
3270	C	AF241410	Hannoun *et al.* (2000)
8290	C	AF241411	Hannoun *et al.* (2000)
975484mut	C	AF330110	
G683-1	C	AF363961	
G683-2	C	AF363962	
G683-3	C	AF363963	
B4-ST1	C	D23680	Horikita *et al.* (1994)
C4-ST2	C	D23681	Horikita *et al.* (1994)
HBAustKW	D	AB048701	Sugauchi *et al.* (2001)
HBAustGC	D	AB048702	Sugauchi *et al.* (2001)
HBAustDF	D	AB048703	Sugauchi *et al.* (2001)
Pat1-*ayw*2	D	AF151735	Gerner *et al.* (1999)
U95551	D	U95551	
Pat-E	D	X65259	
PatD-*ayw*	D	X68292	
Pat4-*ayw*	D	X80924	Karayiannis *et al.* (1995)
Pat6-*ayw*	D	X80925	Karayiannis *et al.* (1995)
Pat3-*ayw*	D	X97849	Alexopoulou *et al.* (1996)
Chimp195	E	AB032431	Takahashi *et al.* (2000)
VNZ8339-2	F	AB036907	Nakano *et al.* (2001)
VNZ8251	F	AB036910	Nakano *et al.* (2001)
VNZ8255	F	AB036911	Nakano *et al.* (2001)
VNZ8323	F	AB036912	Nakano *et al.* (2001)
VNZ8381	F	AB036919	Nakano *et al.* (2001)
VNZ8624	F	AB036920	Nakano *et al.* (2001)
sa4	F	AF223962	Alestig *et al.* (2001)
sa11	F	AF223963	Alestig *et al.* (2001)
sa16	F	AF223965	Alestig *et al.* (2001)
BA5	F	AF043561	Mbayed *et al.* (1998)
BA8	F	AF043573	Mbayed *et al.* (1998)
BA9	F	AF043577	Mbayed *et al.* (1998)
BA10	F	AF043578	Mbayed *et al.* (1998)
FR1	G	AF160501	Stuyver *et al.* (2000)
Mex3c		AF369535	Sanchez *et al.* (2002)
Mex182		AF369536	Sanchez *et al.* (2002)
Mex210		AF369531	Sanchez *et al.* (2002)
Mex350		AF369540	Sanchez *et al.* (2002)
Mex562		AF369543	Sanchez *et al.* (2002)
Mex628		AF369545	Sanchez *et al.* (2002)
Mex637		AF369542	Sanchez *et al.* (2002)

Table 1 (*cont.*)

Strain	Genotype	Accession no.	Reference
ChBASSI		AB046525	Takahashi *et al.* (2000)
Ch926		AF222323	Hu *et al.* (2000)
Chimp2		AF242585	MacDonald *et al.* (2000)
Chimp4		AF242586	MacDonald *et al.* (2000)
Gib824		AJ131568	Grethe *et al.* (2000)
Gib759		AJ131569	Grethe *et al.* (2000)
Gib645		AJ131571	Grethe *et al.* (2000)
Gib160		AJ131572	Grethe *et al.* (2000)
Chimp82		AJ131575	Grethe *et al.* (2000)
Gor97		AJ131567	Grethe *et al.* (2000)
Somad		AF193863	Verschoor *et al.* (2001)
Woolly monkey HBV		NC 001896	Lanford *et al.* (1998)

(Bowyer *et al.*, 1997). Recently, a novel genotype C variant has been found in Australian aborigines (Sugauchi *et al.*, 2001). There has long since been evidence for another genetic variability of HBV by the existence of nine different serological types, called subtypes, of the HBV surface antigen (HBsAg) (Couroucé *et al.*, 1976; Couroucé-Pauty *et al.*, 1983). The molecular basis for the expression of these subtypes has been established (Okamoto *et al.*, 1987b; Norder *et al.*, 1992a). HBV subtypes may belong to either one or several of the different genotypes and hence confer additional heterogeneity within the genotypes (Norder *et al.*, 1992b).

We have reported previously the prevalence of genotype F among HBV-infected patients with Hispanic background in Central America (Arauz-Ruiz *et al.*, 1997a). Phylogenetic analysis of the HBsAg genes of different genotype F strains with varying geographical origin has revealed three different clades within this genotype, where two Nicaraguan strains formed the most divergent clade (Arauz-Ruiz *et al.*, 1997a). Since genotype F is the most divergent genotype, the two Nicaraguan strains might correspond to a distinct genotype at the level of the complete genome. Herein, we report on the complete genomic sequences and phylogenetic analyses of the two Nicaraguan and eight other HBV strains, two of which have shown a high divergence in the small S gene from their respective genotypes.

Methods

■ **Source of HBV DNA.** Serum samples from ten HBeAg-positive chronic HBsAg carriers were used as a source of HBV DNA. Strains 1853Nic, 1889Nic, 1980Nic, 2928Nic, 1116Sal, 70H and 7768H from chronic carriers in Central America have been characterized previously by sequencing the small S gene (Arauz-Ruiz *et al.*, 1997b). All were found to be subtype *adw4*, apart from 1980Nic, which was subtype *ayw4*, and all had been classified as genotype F, although strains 1853Nic and 2928Nic were divergent. HBV strain LAS2523 also encoded subtype *adw4* and differed from the six main HBV genotypes, although it seemed to be more related to genotype F than to any other genotype. This strain derived from a chronic HBV carrier from Los Angeles, USA. Two strains, Z29 and 14/94, were classified as genotype D based on their S genes, although encoding subtype *adw*, and derived from carriers in Spain and Sweden, respectively.

■ **DNA extraction and amplification.** Samples of 5 µl of serum were treated for 2 h with proteinase K in a final volume of 50 µl and heated at 95 °C for 15 min to denature the enzyme, as described previously (Norder *et al.*, 1990). DNA amplification was achieved from 5 µl of digested serum in a 45 µl reaction mixture. The complete HBV genome was amplified by PCR with primers P1, position 1821, and P2, position 1825 (Günther *et al.*, 1995), using the Expand Long Template PCR system (Roche). The product was used to generate two shorter fragments each covering half of the genome. One of the fragments was amplified with primers P1 and K2 (position 637) with a product size of 2056 bp. The other fragment was amplified using primers P2 and K1 (position 252) with a product size of 1591 bp. PCR was performed in a Thermal Controller PTC100 (MJ Research). The annealing temperature used was 60 °C for fragment P1/P2 and 58 °C for fragments P1/K2 and P2/K1. PCR products were separated by agarose gel electrophoresis followed by staining with ethidium bromide.

■ **Sequencing.** The products from two PCR mixtures, P1/K2 or P2/K1, were pooled and purified with the GFX PCR DNA and Gel Band Purification kit (Amersham Pharmacia). DNA quantification was done using GeneQuant (Pharmacia). Sequencing was performed by primer walking in both directions with 20 pmol of template and 4 pmol of each sequencing primer, as described previously (Norder *et al.*, 1994, 1996). Cycle sequencing was performed with a Thermal Controller, PTC100, using the Big Dye Terminator Cycle Sequencing kit, version 2.0 (PE Biosystems). Sequence electrophoresis was carried out with an automated DNA sequencer ABI 3100 (PE Biosystems).

■ **Sequence analyses.** Overlapping fragments obtained were assembled using the SEQMAN software in the DNASTAR package. Each position was confirmed at least twice from different PCR fragments or sequencing runs. The complete genomes were compared with 75 human and 15 non-human primate HBV sequences retrieved from GenBank (Table 1). The strains were distributed as follows: genotypes A, 8; B, 9; C, 25; D, 21; E, 3; F, 9; and G, 1. Alignment was achieved using CLUSTALX software (Thompson *et al.*, 1994) and corrected manually by visual

inspection. Phylogenetic distances were calculated by the DNADIST program in the PHYLIP package, version 3.53c (Felsenstein, 1993). Trees were generated with the UPGMA and neighbour-joining algorithms within the NEIGHBOR program using the woolly monkey HBV genome as the outgroup. Bootstrapping of 500 replicates was done using SEQBOOT and consensus trees were generated by CONSENSE in the PHYLIP package. The obtained trees were visualized with the tree drawing software TREEVIEW (Page, 1996). The MEGA program, version 1.02 (Kumar *et al*., 1994), was used to calculate nucleotide differences between the sequences.

Results

Complete genomes

The genomes of strains 1853Nic, 1889Nic, 2928Nic, 1116Sal, 70H, 7768H and LAS2523 were 3215 nt long, as for genotypes B, C and F, while that of strain 1980Nic comprised only 3161 nt due to deletions in the core gene and the preS region. Both HBV strains from samples Z29 and 14/94 had genomes comprising 3182 nt, similar to other genotype D strains.

To define the magnitudes of inter- and intragenotypic differences to delimit nucleotide divergences within genotypes, pairwise nucleotide comparisons were performed over the complete genomes of 82 HBV strains representing all genotypes (Fig. 1). The mean nucleotide divergences of inter- and intragenotypic comparisons along with corresponding comparisons for the small S genes are shown in Table 2.

Differences between strains within the same genotype ranged from 0·06 to 7·4%. The highest intragenotypic differences, ranging from 5·9 to 7·4%, were found within genotype C, when comparing strains HBAustRC and HBAustSJ with the other genotype C strains. HBAustKW was most divergent within genotype D (5·0–7·1%). For genotype F, the highest differences, 6·3–6·7%, were between the Central American strains and strain Fou. Nucleotide differences between strains belonging to different genotypes ranged from 6·8 to 17·1%, with the lowest differences, 6·8–8·7% (mean 7·3%), between genotype A strains and the genotype D strain PatD-*ayw*.

Strains 1853Nic, 2928Nic and LAS2523 diverged by 0·8–2·5% from each other, 7·2–10·2% (mean 8·1%) from genotype F strains and 13·2–15·7% from other human HBV genotypes. In the histogram for pairwise comparisons, the divergences of these three strains from genotype F strains were within the peak formed by intergenotypic comparisons (Fig. 1). The differences between genotypes F and H were of the same magnitude as the differences between genotypes D and E (7·5–9·6%). Therefore, the three strains mentioned were shown to belong to a different genotype, for which the designation genotype H is proposed. The highest intergenotypic differences were between sequences of strains belonging to genotypes F and H on the one hand and, on the other hand, the other genotypes (12·8–17·1%); this demonstrates a higher level of divergence than other intergenotypic comparisons and explains the gap of 11–12% divergence in the distribution of

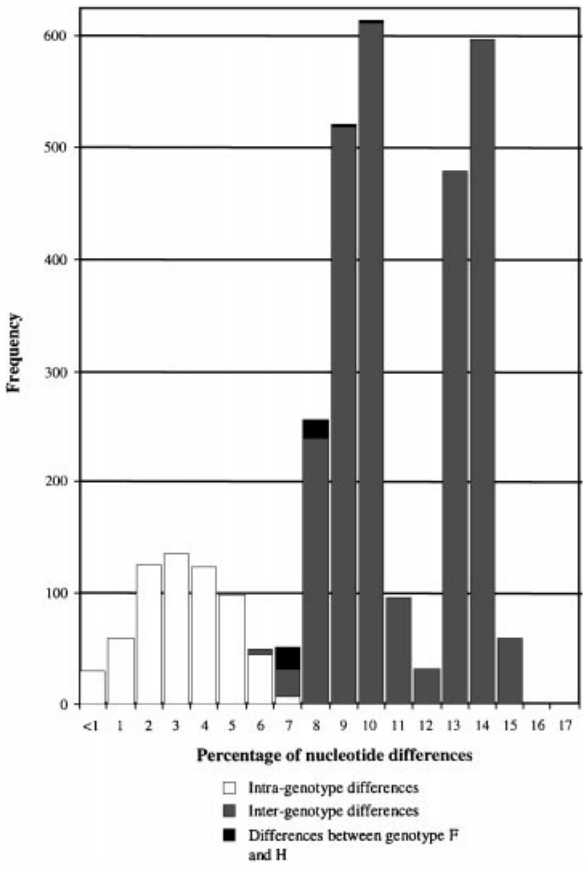

Fig. 1. Distribution of the differences obtained by pairwise comparison of the complete genomes of 82 HBV strains belonging to genotypes A–H. Strains 1114-rec, 6871-rec, 8290 and 3270 were not included in the comparison (Hannoun *et al*., 2000).

nucleotide differences following pairwise comparison (Fig. 1). Differences between strains belonging to genotypes G and other human HBV strains were within the tail of this higher level of differences (12·7–13·8%).

Phylogenetic analysis of the complete genomes of 85 HBV strains derived from humans and 15 strains from non-human primates showed that the genotype H strains were most related to genotype F strains, with which they formed a branch supported by 100% bootstrap (Fig. 2).

Four of the five strains sequenced herein, 1980Nic, 1116Sal, 70H and 7768H, formed a separate clade within genotype F, encompassing the majority of the Central American genotype F strains. These four strains diverged by 0·8–1·4% from each other and 5·1–6·7 % from genotype F strains mainly from South America. 1889Nic belonged to the same clade as the South American genotype F strains and was closely related to a Brazilian strain.

Strains Z29 and 14/94 were within the cluster of genotype D strains and diverged by 3·1% from each other. They showed nucleotide differences of 2·8–4·6% from other genotype D strains and diverged by 7·1–14·0% from strains belonging to other HBV genotypes.

Table 2. Mean per cent nucleotide divergence of the complete genome and the small S gene sequences between 82 human HBV strains belonging to genotypes A–H

Values above the diagonal correspond to comparisons of complete genomes and those below the diagonal correspond to comparisons of the small S genes. Intragenotype divergences are shown in bold; the upper values correspond to comparisons of complete genomes and the lower values to comparisons of the small S genes.

Genotype	A	B	C	D	E	F	G	H
A	**2·3** **1·1**	9·4	9·1	9·7	10·0	14·2	12·3	13·9
B	5·2	**4·1** **1·7**	9·1	10·5	11·0	13·9	13·6	14·1
C	6·1	6·2	**3·6** **2·3**	10·3	10·4	13·7	13·6	13·8
D	4·8	5·7	5·6	**4·0** **1·8**	8·5	14·6	13·1	14·4
E	5·9	5·7	5·5	4·6	**1·2** **1·0**	13·7	11·6	14·0
F	7·8	7·3	7·8	6·6	6·9	**4·0** **1·4**	15·5	8·1
G	4·1	5·7	6·1	4·8	5·1	6·9	**–** **–**	14·5
H	7·9	7·4	8·4	6·9	7·5	3·2	7·4	**1·9** **0·5**

Characterization of different ORFs

The phylogenetic trees obtained after analyses of the individual ORFs and the small S gene showed that the genotype H strains clustered together in a separate branch with those of genotype F, supported by bootstrap values from 98 to 100% (Fig. 3a–f). Also, subgenomic comparisons demonstrated that the genotype F strains from Central America formed a separate cluster within this genotype, apart from strain 1889Nic. Strains Z29 and 14/94 grouped into genotype D (Fig. 3a–f). Nucleotide divergences of different ORFs for the characterized HBV strains as compared to strains belonging to the same and other genotypes are given in Table 3. The number of genotype-specific unique conserved amino acid substitutions for each coding region of genotypes A–H is shown in Table 4.

The preS/S region

The preS region of genotype H strains showed the highest variability as compared to other genotypes. Three conserved amino acid residues, Ala^{8}, Ser^{88} and Pro^{90}, were unique for genotype H in this region. $G1u^{149}$ was shared with the genotype F strains from Central America.

The complete S genes of the four sequenced genotype F strains representing the main Central American cluster diverged by 0·4–1·5% from each other, 2·9–4·8% from other genotype F strains and 11·8–15·3% from strains belonging to other genotypes. In the preS1 region, these four strains shared Asn^{51} with all genotype B and H strains, while other genotype F strains had a Ser in this position. In the preS2 region, these four Central American strains had two residues, Glu^{149} and Leu^{166}, not found in other genotype F strains. The preS regions had the same length as those in other genotype F strains, apart from 1980Nic, which had a deletion of 12 nt (from position 412 to position 423) in the preS2 region.

The small S gene

Phylogenetic analysis of 60 small S gene sequences showed that the genotype H strains grouped together in a separate branch from genotype F supported by 100% bootstrap (Fig. 3c). The presence of Lys^{122}, Leu^{127} and Lys^{160} indicated that they encode subtype *adw4*. Two conserved amino acid residues, Val^{44} and Pro^{45}, within this region were unique. Ten conserved positions were shared only with strains within genotype F, while Ile^{57} was shared with strains belonging to genotypes B and E (Arauz-Ruiz *et al*., 1997b).

Nucleotide divergence between the four sequenced genotype F strains representing the main Central American cluster ranged from 0·3 to 1·2%. They differed by 0·7–3·0% from the other genotype F strains and 5·5–9·2% from the strains of other genotypes. All strains encoded *adw4*, apart from 1980Nic, which encoded *ayw4*. These strains diverged from other genotype F strains with two amino acid substitutions, Glu^{2} instead of Asp^{2} and Thr^{45} instead of Leu^{45} (Arauz-Ruiz *et*

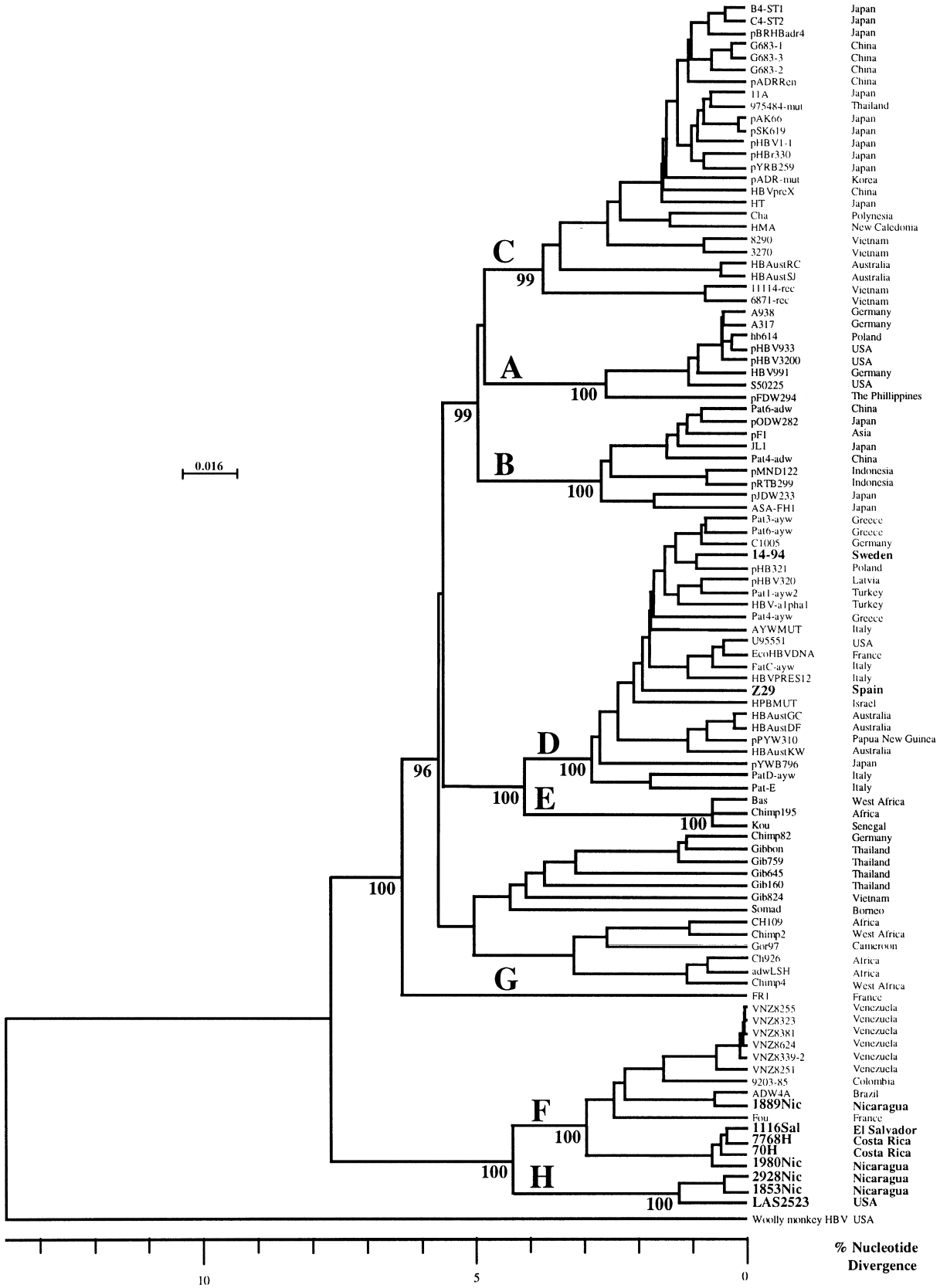

Fig. 2. For legend see facing page.

al., 1997b). Leu^{158} was the only residue unique to all genotype F strains and not present in genotype H.

Both Z29 and 14/94 encoded Lys^{122} and Lys^{160}, characteristic for subtype *adw*. At position 127, which specifies the *w*-type-specific determinant, both strains had a Thr, present in HBV strains encoding the *w*3 specificity (Norder *et al*., 1992b). Two irregular substitutions, Val^{118} and Val^{128}, were shared with another genotype D strain, pHB321.

The precore/core region

The complete C gene was 636 bp long for all strains sequenced herein, apart from 1980Nic, which had a deletion of 42 nt between positions 98 and 139. All three H strains had C^{1858}. Ala^{157} was found as a unique substitution for the three strains. The cytotoxic T lymphocyte epitope in the core region, residues 18–27, was conserved for all strains, apart from Z29, which had Ala^{27}.

The four sequenced genotype F strains representing the main Central American cluster diverged from each other by 0·3–0·9%, 4·4–6·3% from other genotype F strains and 7·4–13·3% from strains within other genotypes. In the precore region, these strains had T^{1858}, whereas other genotype F strains have C^{1858}. They may therefore carry the precore stop mutation G1896 → A. The genotype D strain, Z29, had two translational stop codons at positions 2 and 28 in the precore region.

The X gene

Genotype H strains shared three unique conserved substitutions, Trp^{32}, Ala^{60} and Pro^{102}, in this region. Pro^{34} and Ser^{44} were unique to 1853Nic and 2928Nic. Pro^{92} was unique to 1853Nic, while the other H strains had a unique Thr in this position.

The four sequenced genotype F strains representing the main Central American cluster diverged at the nucleotide level by 0·6–1·3% from one another, 3·2–5·8% from the other genotype F strains and 8·3–15·8% from the other HBV strains. These strains differed by three residues, Ser^{29}, Ser^{31} and Arg^{87}, from other genotype F strains.

The P gene

The P gene was 2532 nt long in the three genotype H strains and in four genotype F strains, encoding a putative protein of 844 aa, as for genotypes B, C and F in general. The P gene of 1980Nic comprised 2520 nt, while this gene of strains Z29 and 14/94 comprised 2499 nt due to the deletions in the core and preS regions.

The amino-terminal part of the polymerase gene, especially the spacer region, was rather variable compared with the carboxy-terminal part. The YMDD motif, from residues 203 to 206, in the reverse transcriptase (RT) region was conserved for all ten strains, as well as the GLY-priming motif, from residues 63 to 65, in the terminal protein (TP). Genotype H had 16 conserved unique substitutions within this region, three in TP, seven in the spacer and three each in RT and the RNase H regions (Fig. 4). The positions immediately next to the catalytic site of the polymerase gene were conserved, apart from a Val^{207} → Leu change in genotypes F and H. Leu^{207} is also present in two genotype C strains.

The four sequenced genotype F strains representing the main Central American cluster differed from the other genotype F strains in 22 positions (Fig. 4).

Discussion

Previously, we have shown using phylogenetic analysis of the S gene that South and Central American HBV strains of genotype F formed three different clades (Arauz-Ruiz *et al*., 1997a). Similar division of genotype F into four clusters, designated I–IV, has been described from Argentina, where the majority of Central American strains are classified as cluster I; our clade formed by two strains from Nicaragua correspond to cluster III (Mbayed *et al*., 2001). Since this clade diverged from the other clades to a degree that they might represent a new genotype, we sequenced and analysed the entire genomes of these two strains and a closely related strain from the USA. Pairwise comparison of their complete genomes with those of other HBV strains revealed that they diverged to the same degree from strains belonging to genotypes A–G as the degree obtained when comparing HBV strains belonging to different genotypes. The three strains could thus be classified into a new genotype. The association to genotype F strains was found in dendrograms based on complete genomes as well as on those based on the different ORFs. The clade formed by the genotype H strains split off from the same main branch as the two clades formed by the genotype F strains in the topology of all phylogenetic trees. The strains of the two genotypes thus shared an internal node, suggesting that genotype F and H strains are descendants of the same ancestral HBV strain.

The majority of differences between genotypes F and H were located by pairwise comparisons within the intergenotypic range (Fig. 1). In contrast, comparisons involving two sequences classified as variants of genotype C with other strains of genotype C were within the range of intragenotypic comparisons. Based on analysis of the small S gene region, they had seemed to belong to a novel genotype distinct from genotype C (Sugauchi *et al*., 2001).

Fig. 2. Dendrogram based on 100 complete HBV genomes: ten from the present study and 90 from GenBank, including 75 derived from humans and 15 from apes. HBV genotypes are designated by the letters A–H. Genomes from this study are indicated in bold. The woolly monkey HBV genome was used as outgroup. Bootstrap values based on 500 replicas are shown at each main branch.

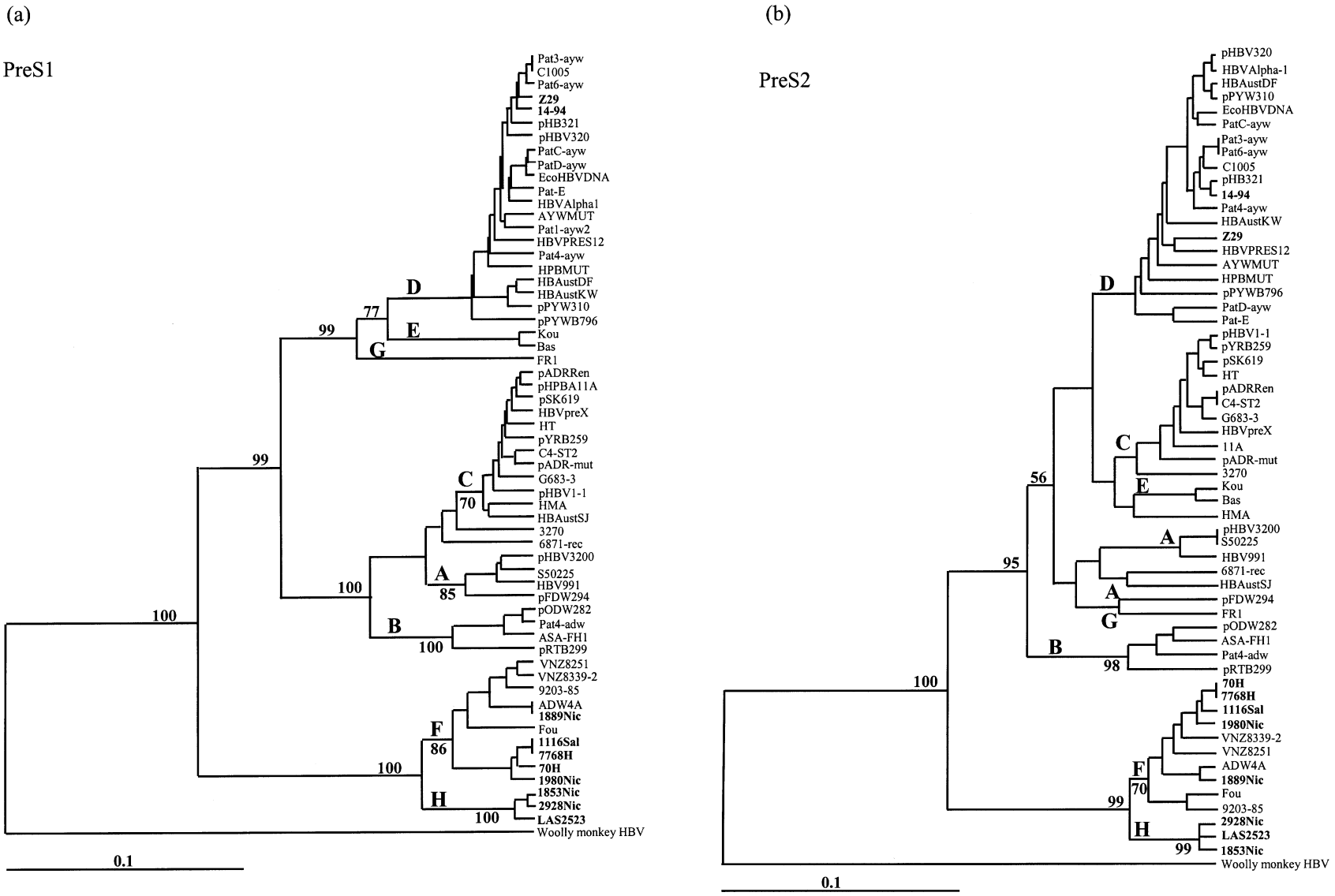

Fig. 3. Dendrograms based on the comparison of ten HBV strains from the present study and 50 strains from GenBank. Regions included in the comparison were: (a) the preS1 gene; (b) the preS2 gene; (c) the small S gene encoding HBsAg; (d) the large S gene, including preS1, preS2 and HBsAg genes; (e) the C gene, including precore and core regions; and (f) the X gene. The same 60 strains are included in (a, b, d–f), whereas (c) includes 14 genotype F strains not used in the other trees. The woolly monkey HBV sequence was used as outgroup. Strains from this study are in bold. Bootstrap values are as for Fig. 2. The scale bar indicates the percentage of nucleotide divergence.

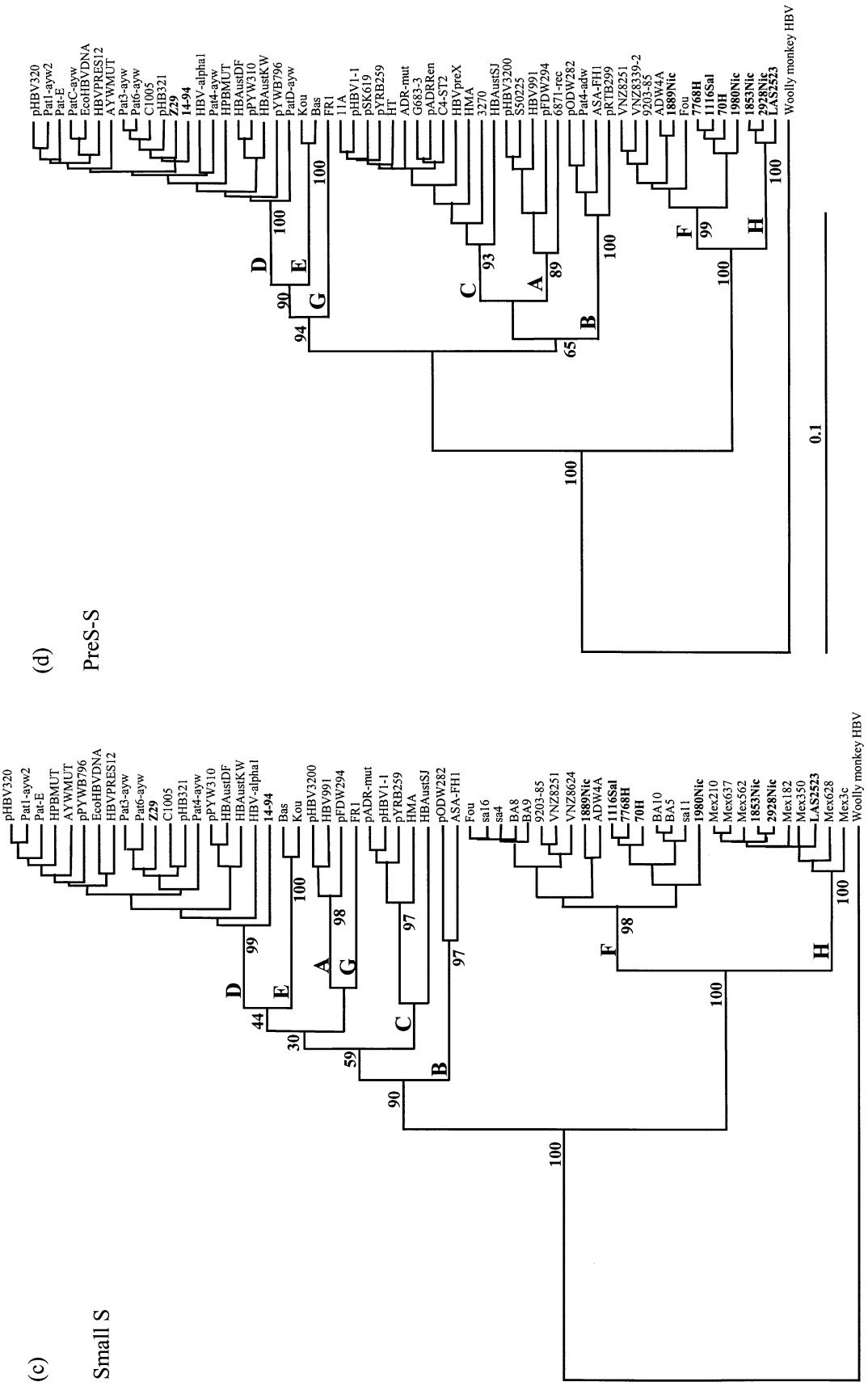

Fig. 3. For legend see opposite.

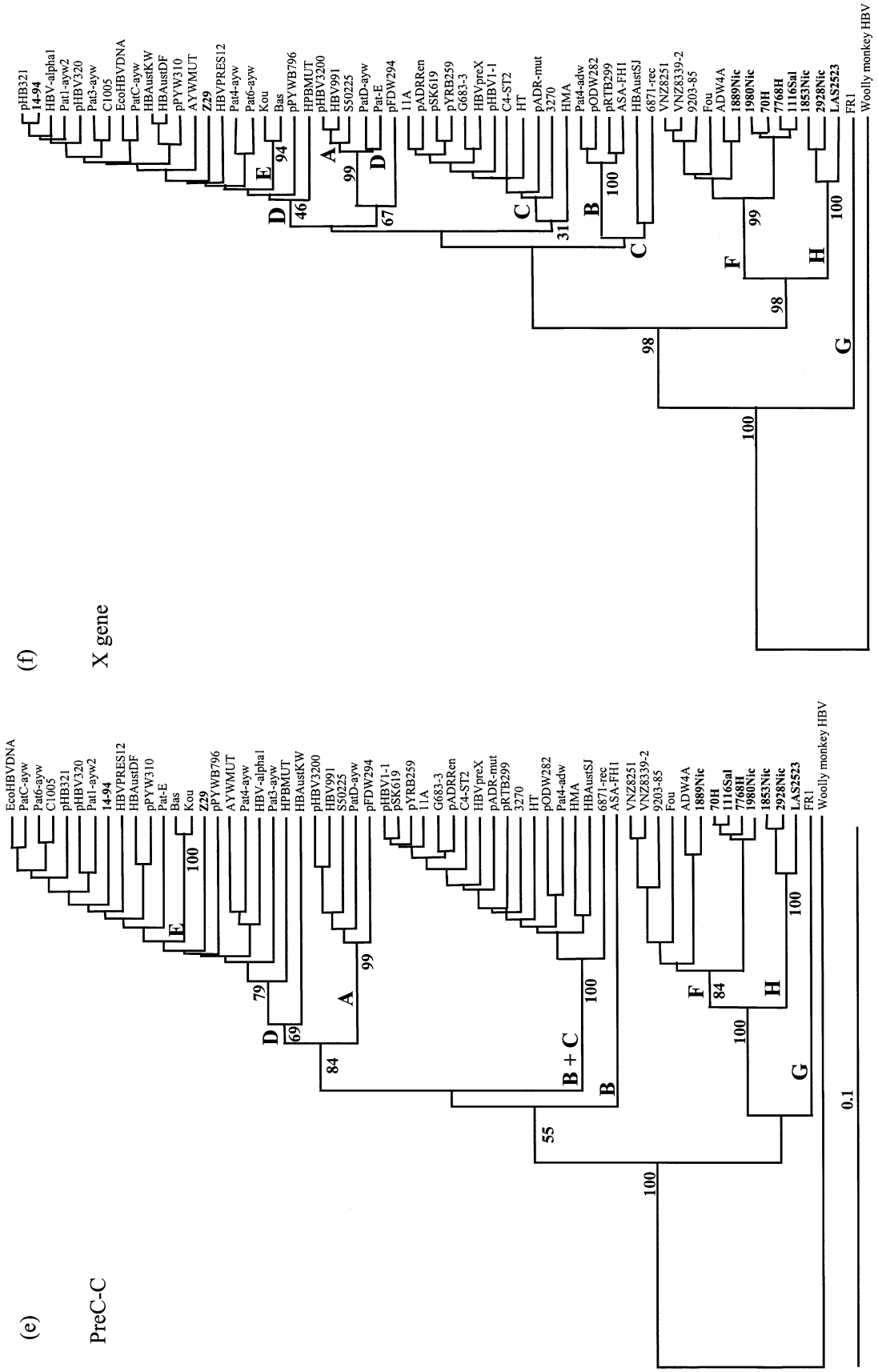

Fig. 3. For legend see p. 2066.

Table 3. Nucleotide divergence of the different genomic regions for five HBV strains sequenced in this study compared to strains in the same and heterologous genotypes

	Genotype H (3 strains)			Genotype D-*adw3* (2 strains)	
Genomic region	H (3 strains) (%)	F (14 strains) (%)	non-F (65 strains) (%)†	D (21 strains) (%)‡	non-D (59 strains) (%)†
PreS	0·7–1·0	6·1–8·5	14·7–22·7	0·9–4·3	8·8–19·6
PreS/S	0·4–1·0	4·7–6·2	11·3–14·6	1·2–4·0	7·1–12·9
HBsAg	0·1–0·6	2·3–4·2	6·3–9·3	0·7–3·8	3·6–7·6
Core*	0·4–2·2	5·5–7·6	9·8–14·4	2·6–7·9	4·5–13·6
X	1·7–4·0	7·0–10·2	8·7–16·2	0·4–6·4	4·0–14·3

* Core region comparisons do not include the insertion of 36 nt in strain FR1 within genotype G.
† Four strains were not included in the comparisons.
‡ The two new sequences of genotype D strains encoding subtype *adw3* were not included.

Table 4. Number of unique conserved amino acid substitutions in the different genes for genotypes A–H based on 82 HBV genomes

	Envelope protein			Core			Polymerase			
ORF	preS1	preS2	HBsAg	preC	Core*	X	TP	Spacer	RT	RNase H
A	4	0	1	1	3	1	5	6	2	0
B	3	2	5	0	0	2	2	10	1	0
C	0	0	3	0	0	0	1	1	0	0
D	3	1	1	0	0	0	1	5	0	0
E	5	0	0	0	0	0	3	9	1	0
F	2	1	2	0	0	2	2	4	0	0
G	4	1	3	0	3	12	2	10	1	1
H	3	0	2	0	1	3	3	7	3	3

* The 12 aa insertion in genotype G is not included.

All genotype D strains described so far have encoded subtype *ayw* (*ayw*2, *ayw*3 or, rarely, *ayw*4). Strains encoding subtype *adw* have so far been found within genotypes A–C and F (Magnius & Norder, 1995). However, two strains encoding subtype *adw*3, Z29 and 14/94, found to cluster with genotype D based on analysis of the small S gene had Thr^{127}, which should define a *w*3 specificity (Norder *et al.*, 1992a). Analysis of the complete genomes of these strains confirmed their classification into genotype D. The presence of a new subtype in genotype D is not unexpected, since subtype heterogeneity has been reported for this genotype (Norder *et al.*, 1993). However, *w*3 specificity, as defined by the presence of Thr^{127}, together with *d* specificity, has not been described before but only in strains encoding *y* specificity. Therefore, it cannot be excluded that *w*3 specificity may be dependent on the presence of Arg^{122} as well as Thr^{127}: *w*1 specificity has been shown to depend on both Arg^{122} and Pro^{127} (Norder *et al.*, 1992a). It is important to keep in mind that subtyping of HBsAg by molecular approaches may not agree with the hypothetical outcome from serological typing with the original antisera if all critical residues for their binding are not defined.

The presence of a double translational stop mutation in codons 2 and 28 of the precore region of Z29 indicates that this strain would not express HBeAg, although this was the case. HBeAg was also present in the serum sample from which a genotype G strain was derived and this strain also had these two precore stop codons (Stuyver *et al.*, 2000). These two substitutions seem to occur naturally in genotype G strains; therefore, an alternative translation strategy seems to be responsible for the presence of HBeAg in the sera of patients infected with this genotype (Kato *et al.*, 2001; Stuyver

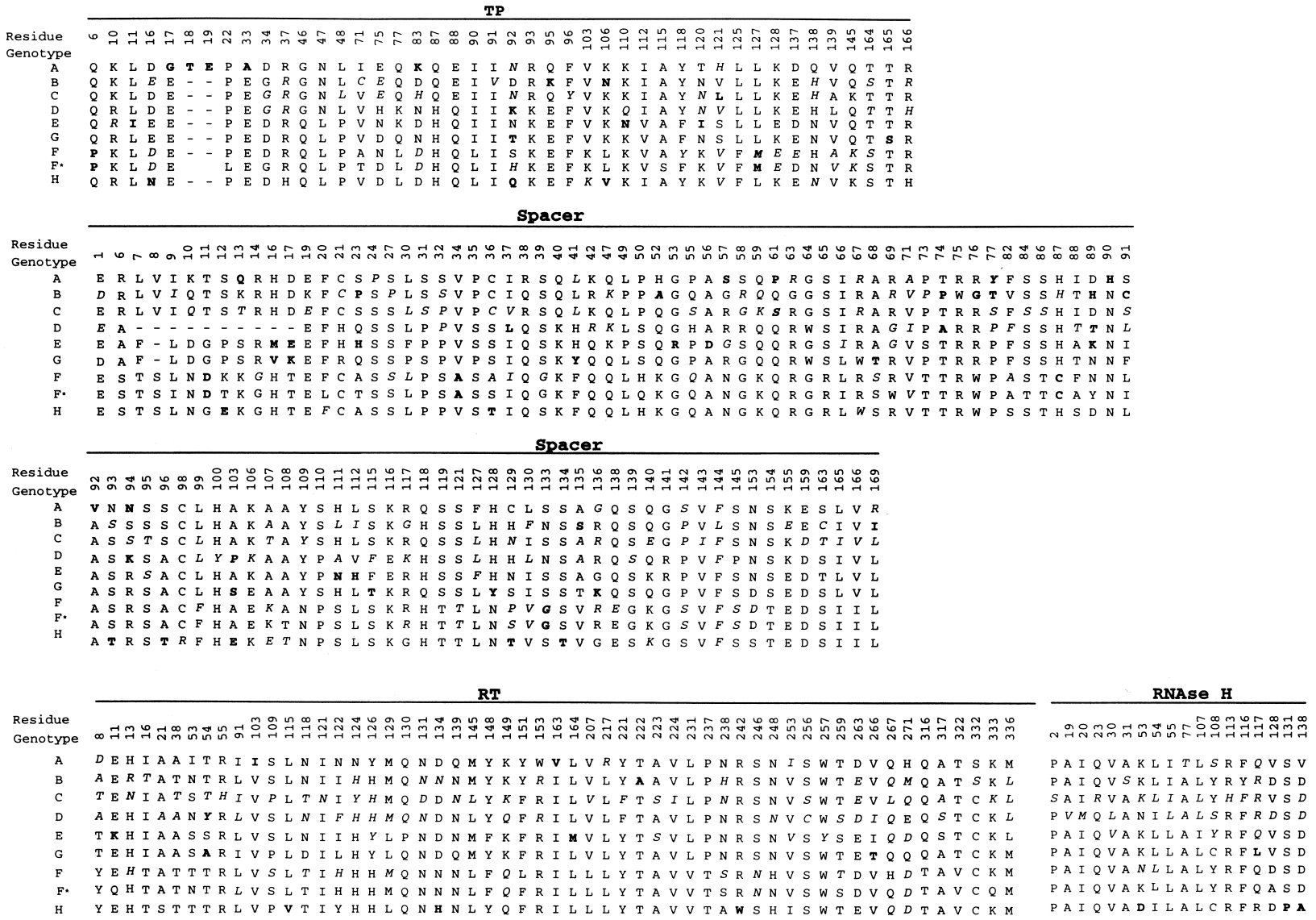

Fig. 4. Amino acid substitutions conserved for genotypes within the primary structure of the P gene based on the comparison of the deduced amino acid sequences from 86 HBV genomes. Bold letters indicate unique substitutions conserved for all compared strains within the genotype. Italic letters indicate the majority substitution if the substitution varies within the genotype. Deletions are represented by dashes. The insertion of 12 residues in the core region of genotype G strains was not included. Alignment and numeration of each genomic region was performed with strain pHBV3200 as reference. The asterisk represents the Central American cluster of genotype F.

et al., 2000). However, it is possible that a heterogeneous virus population exists in these patients, with a predominance of this mutant type but with enough wild-type virus for HBeAg production.

Deletion mutants in the core region are often found in HBV strains from HBeAg-positive patients with chronic hepatitis and are claimed indicative of an ensuing seroconversion to anti-HBe (Günther *et al.*, 1996; Okamoto *et al.*, 1987c; Preikschat *et al.*, 1999b; Tsubota *et al.*, 1998; Wakita *et al.*, 1991). The Nicaraguan strain sequenced herein, 1980Nic, had a 42 nt in-frame deletion resulting in a 14 aa deletion in the amino-terminal region of the deduced core gene product. HBV strains are described, which have deletions of variable sizes in the centre of the core gene between residues 60 and 130, close to the major B cell epitope (Carman *et al.*, 1989; Günther *et al.*, 1996; Marinos *et al.*, 1996; Wakita *et al.*, 1991). The largest of these may also affect the frame of the P gene (Günther *et al.*, 1996). The deletion in 1980Nic was located 25 residues upstream of the B cell epitope and will not affect the polymerase-reading frame. Core deletion mutants always coexist with wild-type HBV genomes, which may provide a helper function for these mutants (Günther *et al.*, 1996; Marinos *et al.*, 1996). Recently, deletion core mutants have been described in genotype F strains from Venezuela (Nakano *et al.*, 2001), where a mixed virus population of wild-type and deletion mutants was reported. None of these deletions is similar to the one in 1980Nic. Depending on their ability to form stable proteins or to assemble into particles, these mutants could contribute to liver cell pathogenesis (Preikschat *et al.*, 1999a).

It has been shown that the genotypes of HBV differ with regard to their association with precore mutants, depending on the nucleotide substitution at position 1858 in the pregenomic encapsidation signal (Li *et al.*, 1993; Rodriguez-Frias *et al.*, 1995). The precore stop mutation, G → A at position 1896, occurs frequently in HBV genotypes with T^{1858}, while this mutation is rare in genotypes with C^{1858}, as with most genotype A strains. The genotype H strains had C^{1858}, suggesting that the precore stop mutation might be rare in this genotype. C^{1858} is also prevalent in the South American genotype F strains, while most genotype F strains from Central America encoded T^{1858}, some of which also carried the precore stop mutation (Arauz-Ruiz *et al.*, 1997a).

It is now evident that genotype F represents the original genotype of the aboriginal populations of the Americas and it has been found in high frequency in several countries in South America, from Argentina to Colombia (Blitz *et al.*, 1998; Casey *et al.*, 1996; Nakano *et al.*, 2001; Naumann *et al.*, 1993; Norder *et al.*, 1993; Telenta *et al.*, 1997), and, more recently, in Central America (Arauz-Ruiz *et al.*, 1997a). All but one of the sequenced strains from Central America formed a separate clade within genotype F, distinct from the one formed mainly by the South American genotype F strains, in phylogenetic trees based on all genetic regions. This seems to reflect different geographical origins of the strains. There are scarce data on the distribution of HBV genotypes among aboriginal populations in North America, including Mexico, USA and Canada, although one genotype F strain has been reported from Alaska (Norder *et al.*, 1993). Genotype A is common in the USA and the genotypes found there mainly reflect the large immigrations from other geographical areas into this country in the past. Recently, however, Sanchez *et al.* (2002) reported the genotypes of 16 HBV strains from Mexico based on HBsAg sequences, seven of which clustered with the genotype H strains in the dendrogram. Thus, it seems that the distribution of genotype H is restricted so far to the Northern part of Latin America, including Central America and Mexico. The genotype H strain from Los Angeles, USA, might be an import from Mexico. However, genotype H may be more widely spread in among USA populations with Hispanic backgrounds, which constitutes 12% of the USA population and is the highest percentage of Hispanics outside Latin America (Therrien & Ramirez, 2000).

Since genotype H is most similar to genotype F, known to be an Amerindian genotype, genotype H has probably split off from genotype F within the New World by early division of the progenitor HBV strains of the first settlers, who arrived to North America around 15 000 years ago (Neel *et al.*, 1994). Analysis of HBV strains from East Asia from where the first Amerindians came may help to clarify the origins and phylogenetic history of genotypes F and H. Since both genotypes F and H are highly divergent from Old World HBV genomes, including those of non-human primates, the branch with genotypes F and H has most likely split off a long time ago from other HBV strains and has a degree of divergence compatible with a cross-species transfer in the past.

This work was supported by Research Grant from Karolinska International Research Program (KIRT)/SAREC.

References

Alestig, E., Hannoun, C., Horal, P. & Lindh, M. (2001). Phylogenetic origin of hepatitis B virus strains with precore C-1858 variant. *Journal of Clinical Microbiology* **39**, 3200–3203.

Alexopoulou, A., Karayiannis, P., Hadziyannis, S. J., Hou, J., Pickering, J., Luo, K. & Thomas, H. C. (1996). Whole genome analysis of hepatitis B virus from four cases of fulminant hepatitis: genetic variability and its potential role in disease pathogenicity. *Journal of Viral Hepatitis* **3**, 173–181.

Arauz-Ruiz, P., Norder, H., Visona, K. A. & Magnius, L. O. (1997a). Genotype F prevails in HBV infected patients of Hispanic origin in Central America and may carry the precore stop mutant. *Journal of Medical Virology* **51**, 305–312.

Arauz-Ruiz, P., Norder, H., Visona, K. A. & Magnius, L. O. (1997b). Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variability of the small S gene. *Journal of Infectious Diseases* **176**, 851–858.

Asahina, Y., Enomoto, N., Ogura, Y., Sakuma, I., Kurosaki, M., Izumi, N., Marumo, F. & Sato, C. (1996). Complete nucleotide sequences of hepatitis B virus genomes associated with epidemic fulminant hepatitis. *Journal of Medical Virology* **48**, 171–178.

Blitz, L., Pujol, F. H., Swenson, P. D., Porto, L., Atencio, R., Araujo, M., Costa, L., Monsalve, D. C., Torres, J. R., Fields, H. A., Lambert, S., Van Geyt, C., Norder, H., Magnius, L. O., Echevarria, J. M. & Stuyver, L. (1998). Antigenic diversity of hepatitis B virus strains of genotype F in Amerindians and other population groups from Venezuela. *Journal of Clinical Microbiology* **36**, 648–651.

Bowyer, S. M., van Staden, L., Kew, M. C. & Sim, J. G. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *Journal of General Virology* **78**, 1719–1729.

Carman, W. F., Jacyna, M. R., Hadziyannis, S., Karayiannis, P., McGarvey, M. J., Makris, A. & Thomas, H. C. (1989). Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* **ii**, 588–591.

Casey, J. L., Niro, G. A., Engle, R. E., Vega, A., Gomez, H., McCarthy, M., Watts, D. M., Hyams, K. C. & Gerin, J. L. (1996). Hepatitis B virus (HBV)/hepatitis D virus (HDV) coinfection in outbreaks of acute hepatitis in the Peruvian Amazon basin: the roles of HDV genotype III and HBV genotype F. *Journal of Infectious Diseases* **174**, 920–926.

Couroucé, A. M., Drouet, J. & Muller, J. Y. (1976). Australia antigen subtypes identification. *Bibliotheca Haematologica* **42**, 89–127.

Couroucé-Pauty, A. M., Plancon, A. & Soulier, J. P. (1983). Distribution of HBsAg subtypes in the world. *Vox Sanguinis* **44**, 197–211.

Felsenstein, J. (1993). PHYLIP: Phylogeny Inference Package, version 3.52c. University of Washington, Seattle, WA, USA.

Gerner, P., Lausch, E., Friedt, M., Tratzmüller, R., Spangenberg, C. & Wirth, S. (1999). Hepatitis B virus core promoter mutations in children with multiple anti-HBe/HBeAg reactivations result in enhanced promoter activity. *Journal of Medical Virology* **59**, 415–423.

Grethe, S., Heckel, J. O., Rietschel, W. & Hufert, F. T. (2000). Molecular epidemiology of hepatitis B virus variants in nonhuman primates. *Journal of Virology* **74**, 5377–5381.

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. *Journal of Virology* **69**, 5437–5444.

Günther, S., Baginski, S., Kissel, H., Reinke, P., Krüger, D. H., Will, H. & Meisel, H. (1996). Accumulation and persistence of hepatitis B virus core gene deletion mutants in renal transplant patients are associated with end-stage liver disease. *Hepatology* **24**, 751–758.

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.

Horikita, M., Itoh, S., Yamamoto, K., Shibayama, T., Tsuda, F. & Okamoto, H. (1994). Differences in the entire nucleotide sequence between hepatitis B virus genomes from carriers positive for antibody to hepatitis B e antigen with and without active disease. *Journal of Medical Virology* **44**, 96–103.

Hu, X., Margolis, H. S., Purcell, R. H., Ebert, J. & Robertson, B. H. (2000). Identification of hepatitis B virus indigenous to chimpanzees. *Proceedings of the National Academy of Sciences, USA* **97**, 1661–1664.

Karayiannis, P., Alexopoulou, A., Hadziyannis, S., Thursz, M., Watts, R., Seito, S. & Thomas, H. C. (1995). Fulminant hepatitis associated with hepatitis B virus e antigen-negative infection: importance of host factors. *Hepatology* **22**, 1628–1634.

Kato, H., Orito, E., Sugauchi, F., Ueda, R., Gish, R. G., Usuda, S., Miyakawa, Y. & Mizokami, M. (2001). Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. *Journal of Virological Methods* **98**, 153–159.

Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer Applications in the Biosciences* **10**, 189–191.

Lanford, R. E., Chavez, D., Brasky, K. M., Burns, R. B., III & Rico-Hesse, R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proceedings of the National Academy of Sciences, USA* **95**, 5757–5761.

Li, J. S., Tong, S. P., Wen, Y. M., Vitvitski, L., Zhang, Q. & Trepo, C. (1993). Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *Journal of Virology* **67**, 5402–5410.

MacDonald, D. M., Holmes, E. C., Lewis, J. C. & Simmonds, P. (2000). Detection of hepatitis B virus infection in wild-born chimpanzees (*Pan troglodytes verus*): phylogenetic relationships with human and other primate genotypes. *Journal of Virology* **74**, 4253–4257.

Magnius, L. O. & Norder, H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* **38**, 24–34.

Marinos, G., Torre, F., Günther, S., Thomas, M. G., Will, H., Williams, R. & Naoumov, N. V. (1996). Hepatitis B virus variants with core gene deletions in the evolution of chronic hepatitis B infection. *Gastroenterology* **111**, 183–192.

Mbayed, V. A., Lopez, J. L., Telenta, P. F., Palacios, G., Badia, I., Ferro, A., Galoppo, C. & Campos, R. (1998). Distribution of hepatitis B virus genotypes in two different pediatric populations from Argentina. *Journal of Clinical Microbiology* **36**, 3362–3365.

Mbayed, V. A., Barbini, L., Lopez, J. L. & Campos, R. H. (2001). Phylogenetic analysis of the hepatitis B virus (HBV) genotype F including Argentine isolates. *Archives of Virology* **146**, 1803–1810.

Nakano, T., Lu, L., Hu, X., Mizokami, M., Orito, E., Shapiro, C., Hadler, S. & Robertson, B. (2001). Characterization of hepatitis B virus genotypes among Yucpa Indians in Venezuela. *Journal of General Virology* **82**, 359–365.

Naumann, H., Schaefer, S., Yoshida, C. F., Gaspar, A. M., Repp, R. & Gerlich, W. H. (1993). Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype *adw4*. *Journal of General Virology* **74**, 1627–1632.

Neel, J. V., Biggar, R. J. & Sukernik, R. I. (1994). Virologic and genetic studies relate Amerind origins to the indigenous people of the Mongolia/Manchuria/southeastern Siberia region. *Proceedings of the National Academy of Sciences, USA* **91**, 10737–10741.

Norder, H., Hammas, B. & Magnius, L. O. (1990). Typing of hepatitis B virus genomes by a simplified polymerase chain reaction. *Journal of Medical Virology* **31**, 215–221.

Norder, H., Couroucé, A. M. & Magnius, L. O. (1992a). Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *Journal of General Virology* **73**, 3141–3145.

Norder, H., Hammas, B., Löfdahl, S., Couroucé, A. M. & Magnius, L. O. (1992b). 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., Hammas, B., Lee, S. D., Bile, K., Couroucé, A. M., Mushahwar, I. K. & Magnius, L. O. (1993). Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural

variations in the primary structure of the surface antigen. *Journal of General Virology* **74**, 1341–1348.

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.

Norder, H., Ebert, J. W., Fields, H. A., Mushahwar, I. K. & Magnius, L. O. (1996). Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* **218**, 214–223.

Okamoto, H., Imai, M., Kametani, M., Nakamura, T. & Mayumi, M. (1987a). Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno–fetal transmission. *Japanese Journal of Experimental Medicine* **57**, 231–236.

Okamoto, H., Imai, M., Tsuda, F., Tanaka, T., Miyakawa, Y. & Mayumi, M. (1987b). Point mutation in the S gene of hepatitis B virus for a *d*/*y* or *w*/*r* subtypic change in two blood donors carrying a surface antigen of compound subtype *adyr* or *adwr*. *Journal of Virology* **61**, 3030–3034.

Okamoto, H., Tsuda, F. & Mayumi, M. (1987c). Defective mutants of hepatitis B virus in the circulation of symptom-free carriers. *Japanese Journal of Experimental Medicine* **57**, 217–221.

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.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Preikschat, P., Borisova, G., Borschukova, O., Dislers, A., Mezule, G., Grens, E., Krüger, D. H., Pumpens, P. & Meisel, H. (1999a). Expression, assembly competence and antigenic properties of hepatitis B virus core gene deletion variants from infected liver cells. *Journal of General Virology* **80**, 1777–1788.

Preikschat, P., Meisel, H., Will, H. & Günther, S. (1999b). Hepatitis B virus genomes from long-term immunosuppressed virus carriers are modified by specific mutations in several regions. *Journal of General Virology* **80**, 2685–2691.

Pult, I., Chouard, T., Wieland, S., Klemenz, R., Yaniv, M. & Blum, H. E. (1997). A hepatitis B virus mutant with a new hepatocyte nuclear factor 1 binding site emerging in transplant-transmitted fulminant hepatitis B. *Hepatology* **25**, 1507–1515.

Rodriguez-Frias, F., Buti, M., Jardi, R., Cotrina, M., Viladomiu, L., Esteban, R. & Guardia, J. (1995). Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. *Hepatology* **22**, 1641–1647.

Sanchez, L. V., Maldnado, M., Bastidas-Ramírez, B. E., Norder, H. & Panduro, A. (2002). Genotypes and S-gene variability of Mexican hepatitis B virus strains. *Journal of Medical Virology* (in press).

Schories, M., Peters, T. & Rasenack, J. (2000). Isolation, characterization and biological significance of hepatitis B virus mutants from serum of a patient with immunologically negative HBV infection. *Journal of Hepatology* **33**, 799–811.

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.

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.

Takahashi, K., Brotman, B., Usuda, S., Mishiro, S. & Prince, A. M. (2000). Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. *Virology* **267**, 58–64.

Telenta, P. F., Poggio, G. P., Lopez, J. L., Gonzalez, J., Lemberg, A. & Campos, R. H. (1997). Increased prevalence of genotype F hepatitis B virus isolates in Buenos Aires, Argentina. *Journal of Clinical Microbiology* **35**, 1873–1875.

Therrien, M. & Ramirez, R. (2000). The Hispanic population in the United States: March 2000, pp. 520–535. Census Bureau, Washington DC, USA.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Tsubota, A., Kumada, H., Takaki, K., Chayama, K., Kobayashi, M., Suzuki, Y., Saitoh, S., Arase, Y., Murashima, N. & Ikeda, K. (1998). Deletions in the hepatitis B virus core gene may influence the clinical outcome in hepatitis B e antigen-positive asymptomatic healthy carriers. *Journal of Medical Virology* **56**, 287–293.

Verschoor, E. J., Warren, K. S., Langenhuijzen, S., Heriyanto, Swan, R. A. & Heeney, J. L. (2001). Analysis of two genomic variants of orang-utan hepadnavirus and their relationship to other primate hepatitis B-like viruses. *Journal of General Virology* **82**, 893–897.

Wakita, T., Kakumu, S., Shibata, M., Yoshioka, K., Ito, Y., Shinagawa, T., Ishikawa, T., Takayanagi, M. & Morishima, T. (1991). Detection of pre-C and core region mutants of hepatitis B virus in chronic hepatitis B virus carriers. *Journal of Clinical Investigation* **88**, 1793–1801.

Yang, Z., Lauder, I. J. & Lin, H. J. (1995). Molecular evolution of the hepatitis B virus genome. *Journal of Molecular Evolution* **41**, 587–596.

Received 11 December 2001; Accepted 21 March 2002