Journal of General Virology (2004), **85**, 1211–1220 DOI 10.1099/vir.0.19749-0

Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa

Gerald C. Kimbi, Anna Kramvis and Michael C. Kew

MRC/CANSA/University Molecular Hepatology Research Unit and Department of Medicine, University of the Witwatersrand, Parktown 2193, Johannesburg, South Africa

Correspondence
Anna Kramvis
kramvisa@medicine.wits.ac.za

Phylogenetic analysis of hepatitis B virus (HBV) has led to its classification into eight genotypes, A to H. The dominant genotype in South Africa is genotype A, which consists of two subgenotypes, A1 and A2. Subgenotype A1 (previously subgroup A′) predominates over subgenotype A2 (previously subgroup A minus A′). The complete genome of HBV isolated from 18 asymptomatic carriers of the virus and five acute hepatitis B patients was amplified; the resulting amplicons were cloned and sequenced. All acute hepatitis isolates belonged to subgenotype A1 and had no distinguishing mutations relative to the isolates from asymptomatic carriers, which had a distribution of ten subgenotype A1, two subgenotype A2 and six genotype D. The presence of the previously described amino acid residues that distinguish subgenotype A1 (subgroup A′) from the remainder of genotype A in the S and polymerase genes was confirmed. Moreover, the large number of subgenotype A1 isolates sequenced allowed identification in the other open reading frames of additional nucleotide and amino acid changes that are characteristic of subgenotype A1. In particular, nucleotide mutations at positions 1809–1812 that alter the Kozak sequence of the precore/core open reading frame, and A^{1888} in the precore region, were found exclusively in subgenotype A1 isolates. Unique sequence alterations of the transcriptional regulatory elements were also found in subgenotype A1 isolates. The mean nucleotide divergence of subgenotype A1 was greater than that of subgenotype A2, suggesting that this subgenotype has been endemic for a longer time in the South African black population than had subgenotype A2.

Received 22 October 2003
Accepted 19 January 2004

INTRODUCTION

Hepatitis B virus (HBV), a DNA virus, is a member of the family *Hepadnaviridae*. Infection with HBV is a public health problem of worldwide importance, with some 387 million people being chronically infected (WHO, 2002). In sub-Saharan Africa the virus is hyperendemic, with a carrier rate of 5 to 20 % (Mphahlele *et al.*, 2002). HBV infection causes a spectrum of liver diseases, including subclinical, acute self-limited, and fulminant hepatitis, an asymptomatic carrier (ASC) state, chronic hepatitis progressing to cirrhosis, and hepatocellular carcinoma.

Hepadnaviruses have an unusual mechanism of viral DNA replication involving reverse transcription of pregenomic RNA by the virus-encoded polymerase (Nassal & Schaller, 1996). The virus polymerase lacks proofreading activity, and sequence heterogeneity is a feature of HBV. Phylogenetic analysis of HBV full-length genomes has led to the classification of HBV into eight genotypes. The separate genotypes are arbitrarily defined by an intergroup divergence in the complete HBV genome sequence of more than 8 % (Norder *et al.*, 1992a; Okamoto *et al.*, 1988) and at the level of the S gene of more than 4 % (Norder *et al.*, 1992b). Early studies enabled the identification of four genotypes, A to D (Okamoto *et al.*, 1988), with the genotypes E, F (Norder *et al.*, 1992b, 1994), G (Stuyver *et al.*, 2000) and H (Arauz-Ruiz *et al.*, 2002) being identified later.

The eight genotypes show a distinctive geographical distribution. Genotype A is prevalent in northwestern Europe, North America and Africa (Norder *et al.*, 1993). Genotypes B and C are characteristic of Asia (Okamoto *et al.*, 1988), whereas genotype D has a worldwide distribution but predominates in the Mediterranean area. Genotype E is found in Africans (Odemuyiwa *et al.*, 2001), genotype F in the aboriginal populations of South America (Arauz-Ruiz *et al.*, 1997; Norder *et al.*, 1993) and genotype H is confined to the Amerindian populations of Central America (Arauz-Ruiz *et al.*, 2002). To date, the isolation of genotype G has been limited to HBV carriers in France and Georgia, USA (Stuyver *et al.*, 2000) and Germany (Vieth *et al.*, 2002).

The accession numbers of hepatitis B virus isolates sequenced in this study have been deposited in GenBank/EMBL/DDBJ as AY233274–AY233296.

Genotypes A and D coexist in southern Africa, with genotype A predominating. Furthermore, a unique segment of genotype A, subgroup A′ (renamed subgenotype A1) has been identified in this region (Bowyer *et al.*, 1997; Kramvis *et al.*, 2002). This subgroup diverges from subgroup A minus A′ (renamed subgenotype A2). Because subgenotype A1 is widespread in southern Africa it is important to study the influence of this subgenotype on virus replication and disease outcome (Mayerat *et al.*, 1999). Our previous study on subgroup A′ (subgenotype A1) was limited to isolates from five patients with fulminant hepatitis and one with acute hepatitis (AH) (Kramvis *et al.*, 2002). Therefore, the aim of the present study was to carry out full genome analysis on HBV isolates from a larger number of AH patients and from ASCs of the virus. A comparison of these sequences with HBV isolates from various parts of the world and with various disease conditions may help us in understanding the pathogenesis of HBV-induced disease.

METHODS

Subjects and serum samples. Serum samples were collected from 275 hepatitis B surface antigen (HBsAg)-positive southern African blacks: 260 were ASCs and 15 AH patients. The ASCs were randomly selected, unrelated factory workers and labourers from the Gauteng Province. The AH patients were treated at the Johannesburg Academic Hospital as sporadic cases during recent years. These samples were stored at −70 °C until analysed. Commercially available kits (Abbott) were used to detect HBV markers in the serum. The study was approved by the Human Ethics Committee of the University of the Witwatersrand and informed consent was obtained from all subjects.

DNA extraction, amplification, cloning and sequencing. Total DNA was extracted from the sera using the QIAamp DNA Mini Kit (Qiagen), according to the manufacturer's instructions. When the virus concentration was high enough, the complete genome of the virus was amplified using a single amplification method (Gunther *et al.*, 1995) using primers P1 (5′-TTTTTCACCTCTGCCTAATCA-3′) (1821–1843 from *Eco*RI site) and P2 (5′-AAAAAGTTGCAT-GRTGMTGG-3′) (1825–1801 from *Eco*RI site). However, when the virus load was too low for complete genome amplification using single-round PCR, a modification of two subgenomic PCRs was used (Takahashi *et al.*, 1998); this involved the amplification of two overlapping fragments of HBV, fragment A (1·35 kb) and fragment B (2·2 kb) (Table 1). This PCR was designed so that the overlap occurred over the variable regions of the S and precore/X genes, which would allow us to conclude that the amplified DNA was from a single genome, when overlapping regions were identical.

The reaction mix for the amplification consisted of 2·25 µl 10 × Ex Taq buffer with 20 mM $MgCl_2$, 2 µl 2·5 mM dNTP mix, 1·25 µl each of the appropriate primers (Table 1), 2·5 µl DNA, made up to 22·5 µl with water. The enzyme mix was made up of 1·875 µl water, 0·25 µl 10 × Ex Taq buffer and 0·375 µl TaKaRa Ex Taq polymerase. The 22·5 µl reaction mix was preheated to 94 °C for 2 min and 2·25 µl TaKaRa Ex Taq enzyme mix was added at the first annealing step. This was followed by 40 cycles of amplification with the cycling profile shown in Table 1.

Amplicons were cloned into a pPCR-Script Amp SK+ vector (Stratagene) according to the protocol provided by the manufacturer. The positive clones containing the correct size amplicons were prepared for direct sequencing using the BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and sequenced on an Applied Biosystems 377 DNA automated sequencer using vector-specific primers T3 (5′-AATTAACCCTCACTAAAG-GG-3′) and T7 (5′-GTAATACGACTCACTATAGGGC-3′) as well as HBV-specific primers (Owiredu *et al.*, 2001). All sequences were analysed in both the forward and reverse directions.

Phylogenetic analyses. Complete HBV genomes sequences were compared with corresponding sequences of HBV from GenBank. Multiple sequence alignments were carried out using Dambe (Xia, 2000). The alignments were edited manually in GeneDoc (Nicholas & Nicholas, 1997) and fed into PHYLIP (phylogeny inference package) version 3.5c (Felsenstein, 1995). DNAML (maximum-likelihood) alone and DNADIST consecutively with NEIGHBOR (neighbour-joining) were used to generate dendrograms. SEQBOOT, DNADIST and NEIGHBOR were used for bootstrapping of 1000 datasets. CONSENSE was used to compute a consensus tree. Trees were visualized using TreeView Win 32 software program (Page, 1996).

RESULTS

Amplification, cloning and sequencing

Using subgenomic nested PCR assays, 223 of 260 (86 %) ASC sera and all 15 (100 %) AH patient sera were HBV DNA positive. The complete genomes of HBV isolates from 24 ASCs were successfully amplified, 20 using the single round Gunther method (Gunther *et al.*, 1995) and four a modification of the method described by Takahashi *et al.* (2000). The complete genomes of five HBV isolates from

Table 1. Oligonucleotide primers and polymerase chain reaction cycling profiles

(+) sense (-) anti-sense.

Fragment	Primer	Position*	Sequence	Denaturation	Annealing	Extension	Size†
A	455(+)	455-474	5′CAAGGTATGTTGCCCGTTTG3″	94 °C 30 sec	62 °C 30 sec	72 °C 90 sec	1345
	1800(-)	1800-1773	5′AGACCAATTTATGCCTACAGCCTCCTA3′				
B	1687(+)	1687-1708	5′CGACCGACCTTGAGGCATAC3′	94 °C 30 sec	63 °C 30 sec	72 °C 120 sec	2198
	685(-)	704-685	5′CGAACCACTGAACAAATGGC3′				

*Denotes the nucleotide position of hepatitis B virus *adw* genome (GenBank accession #V00866) where the *Eco*RI cleavage site is position 1.
†Size of the amplicons in base pairs.

AH patients were successfully amplified using the method of Takahashi *et al.* (2000). Of these, 18 amplicons from ASCs and the five from AH patients were successfully cloned and sequenced. These sequences have been deposited in GenBank/EMBL/DDBJ databases as AY233274–AY233296. Subgenomic PCR and direct sequencing of the amplicons confirmed that the clones sequenced were representative of the major HBV strain in the serum.

Phylogenetic analysis

The length of the complete genomes of all South African (SA) genotype A isolates sequenced in the present study was 3221 bp. The serological subtype of all SA genotype A isolates, except AY233288, was deduced from the sequence to be *adw2*. Isolate AY233288 belonged to serological subtype *ayw2*. The complete genome sequences of the 23 HBV isolates sequenced were aligned with 32 sequences from GenBank and phylogenetic analysis was carried out (Fig. 1). All AH isolates belonged to subgenotype A1 and had no distinctive mutations relative to the isolates from the ASCs. The HBV genotype distribution among the 18 isolates from ASCs was ten subgenotype A1, two subgenotype A2 and six genotype D. The isolates clustered in the same positions when phylogenetic analysis of the individual open reading frames (ORFs) was performed. There was no unique clustering of the isolates from ASCs when compared to those from AH patients.

The complete genomes of subgenotype A1 had a mean nucleotide divergence from subgenotype A2 of >4% (Table 2), validating the existence of a separate subgroup of genotype A. HBV isolates belonging to subgenotype A1 had a higher intragroup divergence than isolates belonging to subgenotype A2 (Table 2).

Comparison of amino acid sequences of subgenotype A1 to those of subgenotype A2 and other genotypes

Fig. 2. provides a comparison of the translated sequences of the 15 subgenotype A1 and two subgenotype A2 isolates, sequenced in the present study, to sequences obtained from GenBank. The amino acids found only in subgenotype A1 are shaded in grey, whereas those that are found in subgenotype A1 and in other non-A genotypes, but not in subgenotype A2, are shown in bold. Confirming our previous analyses, amino acids Gln^{54}, Val^{74}, Ala^{86} and Val^{91} in the pre-S1 region; Leu^{32} in pre-S2 and Thr^{236}, Gly^{268}, Tyr^{269}, Gln^{334}, Lys^{338} in the spacer of the polymerase gene were unique to subgenotype A1 (Bowyer *et al.*, 1997; Kramvis *et al.*, 2002). In addition, in the present study, we identified residues Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146}, Ser^{147} in the X ORF to be characteristic of subgenotype A1. Residues His^{182} and Ser^{251} in the polymerase ORF are shared by subgenotype A1 and genotype E, the genotype found only in Africa.

By corollary, subgenotype A2 also has signature amino acids that distinguish it from subgenotype A1 and all other genotypes. These are circled in Fig. 2. They are Ala^{54}, Ser^{89}, Thr^{90} and Ile^{91} in the pre-S1; Ala^{47} in the pre-S2; Val^{209} in the S region. In the polymerase region, subgenotype A2 had the following unique amino acids: Gly^{18}, Ala^{33} and Thr^{120} in the priming region and Val^{271}, Asp^{273}, Cys^{308} and Arg^{348} in the spacer of the polymerase.

Comparison of the nucleotide sequences of *cis*-acting elements of subgenotype A1 with those of subgenotype A2 and other genotypes

The sequences of the *cis*-acting elements of the 15 subgenotype A1 and 2 subgenotype A2 isolates sequenced in the present study were compared with sequences obtained from GenBank (Fig. 3). Nucleotides found in subgenotype A1 but not A2 are shown in bold, and those that are found predominantly in subgenotype A1 but not other genotypes are shown in bold and shaded. Table 3 summarizes the mutations in the *cis*-acting regulatory elements characteristic of subgenotype A1 and the functional elements that are affected by these mutations.

DISCUSSION

In our previous study we presented a phylogenetic analysis of four complete and three pre-S1/S2/S gene sequences of subgenotype A1 (previously called subgroup A′) HBV isolates from fulminant and AH patients from SA (Kramvis *et al.*, 2002). Here we extend the study to isolates from both ASCs and additional AH patients with the analysis of 15 complete genomes belonging to subgenotype A1 and two to subgenotype A2 (Fig. 1). This allows confirmation of our previous findings regarding subgenotype A1 (Bowyer *et al.*, 1997; Kramvis *et al.*, 2002), as well as a more comprehensive analysis of subgenotype A1 from Africa and identification of features unique to this subgenotype.

All isolates belonging to subgenotype A1 sequenced in this study had a genome length of 3221 bp and did not have the 21 bp deletion, which we had found previously in the isolates from fulminant hepatitis patients (Kramvis *et al.*, 2002). This deletion was also absent from subgenotype A1 isolates from chronic carriers from Malawi (Sugauchi *et al.*, 2003).

The amino acids differentiating the subgenotypes of genotype A from each other and from other genotypes were concentrated in the pre-S1 region overlapping the spacer of the polymerase (Fig. 2). The sequence of the pre-S1 region has been shown to be well conserved within a given HBV subtype (Uy *et al.*, 1992) and this region may play a role in the attachment of HBV to hepatocytes (Neurath *et al.*, 1986; Pontisso *et al.*, 1989). Therefore, it is possible that the molecular evolution of the pre-S1 sequence is constrained by the host population (Kramvis *et al.*, 2002). It is of interest to note that Gln^{54} and Val^{91} in the pre-S1 region and Thr^{236} in the spacer of the polymerase, unique to subgenotype A1 isolates, are also found in the aberrant

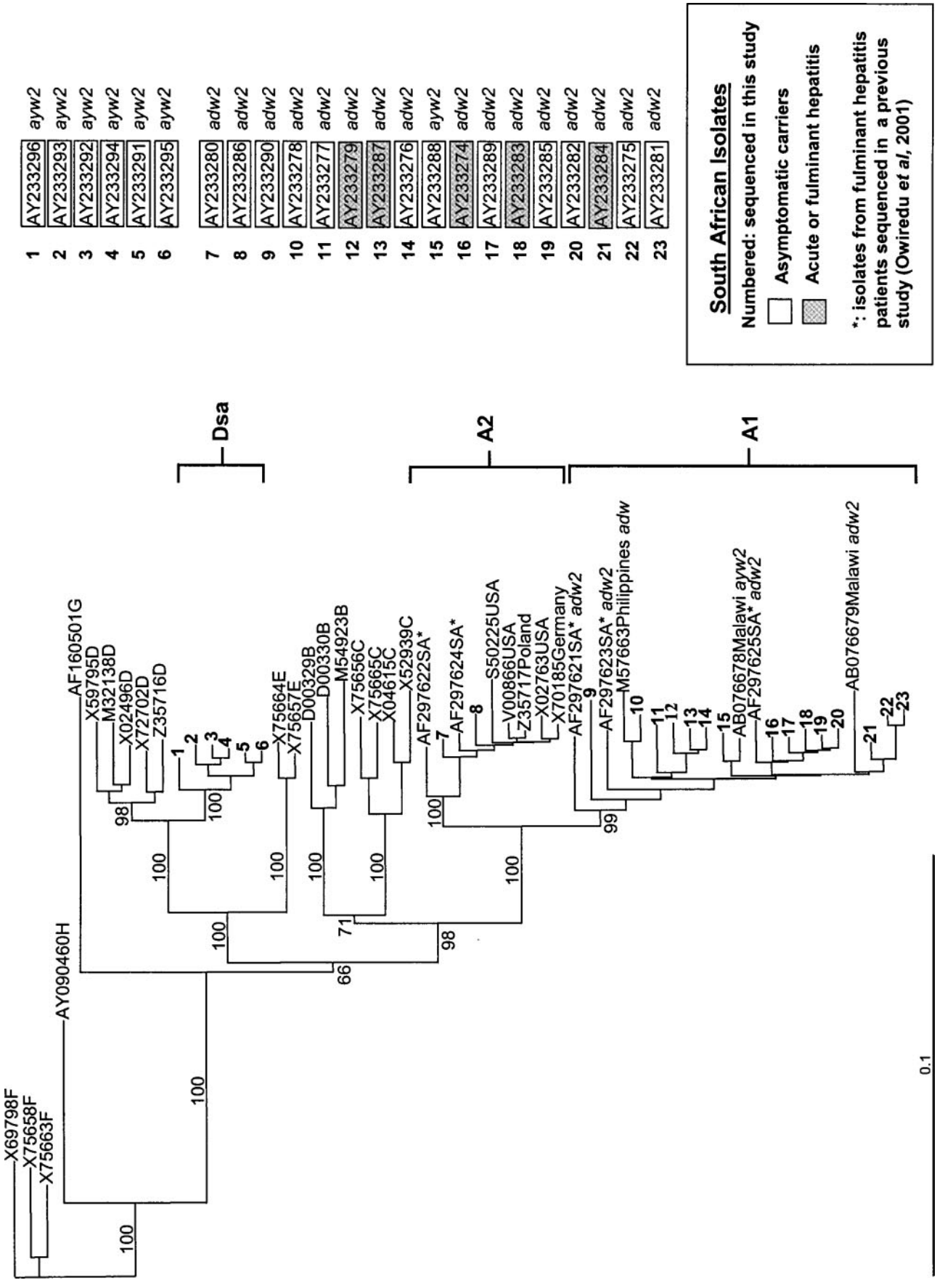
South African Isolates
Numbered: sequenced in this study
Asymptomatic carriers
Acute or fulminant hepatitis
*: isolates from fulminant hepatitis patients sequenced in a previous study (Owiredu et al, 2001)
1 AY233296 ayw2
2 AY233293 ayw2
3 AY233292 ayw2
4 AY233294 ayw2
5 AY233291 ayw2
6 AY233295 ayw2
7 AY233280 adw2
8 AY233286 adw2
9 AY233290 adw2
10 AY233278 adw2
11 AY233277 adw2
12 AY233279 adw2
13 AY233287 adw2
14 AY233276 adw2
15 AY233288 ayw2
16 AY233274 adw2
17 AY233289 adw2
18 AY233283 adw2
19 AY233285 adw2
20 AY233282 adw2
21 AY233284 adw2
22 AY233275 adw2
23 AY233281 adw2
Dsa
A2
A1
X69798F
X75658F
X75663F
AY090460H
AF160501G
X59795D
M32138D
X02496D
X72702D
Z35716D
X75664E
X75657E
D00329B
D00330B
M54923B
X75656C
X75665C
X04615C
X52939C
AF297622SA*
AF297624SA*
S50225USA
V00866USA
Z35717Poland
X02763USA
X70185Germany
AF297621SA* adw2
AF297623SA* adw2
M57663Philippines adw
AB076678Malawi ayw2
AF297625SA* adw2
AB076679Malawi adw2
0.1

Table 2. Mean nucleotide divergence (%) of complete genome and individual open reading frame (ORF) sequences of HBV obtained using DAMBE*

	Intragroup			Intergroup
	Subgenotype A1	Subgenotype A2	Genotype A	A1 vs A2
Complete genome	2·53±1·00 (0·63-4·69)†	1·39±0·87 (0·48-3·49)	3·36±1·47 (0·48-3·49)	4·74±0·56 (3·50-6·72)
polymerase	2·52±0·95 (0·61-4·44)	1·23±0·70 (0·41-2·80)	3·35±1·48 (0·41-6·42)	4·77±0·49 (3·63-6·42)
Pre-S1/S2	3·39±2·07 (0·43-9·07)	1·45±1·02 (0·43-7·68)	4·72±2·36 (0·43-9·07)	6·99±1·02 (0·41-9·07)
HBsAg	1·21±0·56 (0·29-3·67)	0·84±0·62 (0·15-2·35)	1·53±0·74 (0·15-3·67)	2·06±0·59 (1·03-3·67)
Precore/core	2·40±1·19 (0·31-5·75)	1·75±1·67 (0·00-5·91)	3·26±1·65 (0·00-8·24)	4·58±1·08 (2·64-8·24)
X	1·78±1·31 (0·00-6·45)	1·55±0·71 (0·00-3·01)	2·38±1·26 (0·57-6·45)	3·23±0·69 (1·72-4·95)

*http://web.hku.hk/~xxia/software/software.htm. The sequences compared are those included in figure 1 [21 subgenotype A1 and 9 subgenotype A2].

†The mean nucleotide divergence (%) ± standard deviation and the range in parentheses.

genotype A HBV recognized in Vietnam, and it has been suggested that this aberrant genotype may be a link between the European/African A and the Asian B and C genotypes (Hannoun *et al.*, 2000). Val^{91} of the pre-S1 gene that is characteristic of subgenotype A1 is also found in gibbon (Grethe *et al.*, 2000; Norder *et al.*, 1996) and orang-utan (Verschoor *et al.*, 2001) hepadnavirus isolates.

The subgenotype A1 unique amino acids Gln^{334} and Lys^{338} are found in the fingers of the HBV polymerase within the DNA-binding cleft that is positively charged (Das *et al.*, 2001). Gly^{334} is uncharged and replaces basic Lys in subgenotype A2 and acidic Glu in other genotypes. On the other hand, basic Lys^{338} replaces Glu and Asp, both of which are acidic and found in subgenotype A2 and other genotypes, respectively. The change in the charge caused by the alternate amino acids within this region could possibly affect the binding of the DNA to the polymerase and influence reverse transcription. Subgenotype A1, which is the subgenotype prevalent in southern Africa is associated with low HBV DNA levels (Kramvis *et al.*, 1997).

The following signature amino acid motif, 'Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146}, Ser^{147}', was recognized in the X region of ten SA subgenotype A1 isolates and in subgenotype A1 isolates from the Philippines and Malawi (Fig. 2). All the amino acids changes in subgenotype A1 versus A2 and the other genotypes were Pro to Ser and Ser to Ala or vice versa. These changes were found in regions of the HBx protein that are not functionally active in transactivation (Arii *et al.*, 1992). The Ser^{146}, Ser^{147} in the X ORF are a result of mutations at nucleotide position 1809 and 1812 and have an effect on the overlapping Kozak sequence preceding the precore/core start codon.

Double or triple point mutations at positions 1809–1812 were found only in subgenotype A1 isolates and not in subgenotype A2 or other genotypes (Fig. 3). In a previous study, we reported that 80 % of SA HBV strains harbour similar mutations immediately upstream of the precore AUG codon (Baptista *et al.*, 1999), which might impair HBeAg expression as a result of suboptimal translational initiation (Kozak, 1986, 1987; Kramvis & Kew, 1999). We tested this hypothesis using site-directed mutagenesis and transfection experiments and showed that hepatitis B e antigen expression was severely impaired by the $1809^{T}1811^{T}1812^{T}$ and $1809^{T}1811^{C}1812^{T}$ triple mutations, and moderately reduced by the $1809^{T}1812^{T}$ and $1809^{A}1812^{T}$ double mutations (Ahn *et al.*, 2003). The effect of the double mutations on hepatitis B e antigen expression was comparable with that of the common core promoter mutations ($1762^{T}/1764^{A}$) and independent of HBx expression (Ahn *et al.*, 2003). These mutations are not a result of an adaptive change under immune pressure because they are found in HBV isolates obtained from children and in acute hepatitis patients (Ahn *et al.*, 2003). These mutations have previously been reported to occur only occasionally in other regions of the world (Estacio *et al.*, 1988, Kidd-Ljunggren *et al.*, 1995, Laskus *et al.*, 1994), supporting our observation that they are characteristic of subgenotype A1 isolates.

Regulatory *cis*-acting elements are embedded within the protein coding ORFs of HBV. Therefore any changes in

Fig. 1. Phylogenetic relationship of 23 HBV isolates from South Africa [numbered 1 to 23] to full-length sequences of other HBV isolates obtained from GenBank established using neighbour-joining. Bootstrap statistical analysis was performed using 1 000 datasets and the numbers on the nodes indicate the percentage of occurrences. Each sequence obtained from GenBank is designated by its accession number. A1: subgenotype A1 (previously subgroup A′ (Bowyer *et al.*, 1997; Kramvis *et al.*, 2002); A2: subgenotype A2 (previously subgroup A-A′; Bowyer *et al.*, 1997; Kramvis *et al.*, 2002); Dsa: clade containing SA genotype D isolates.

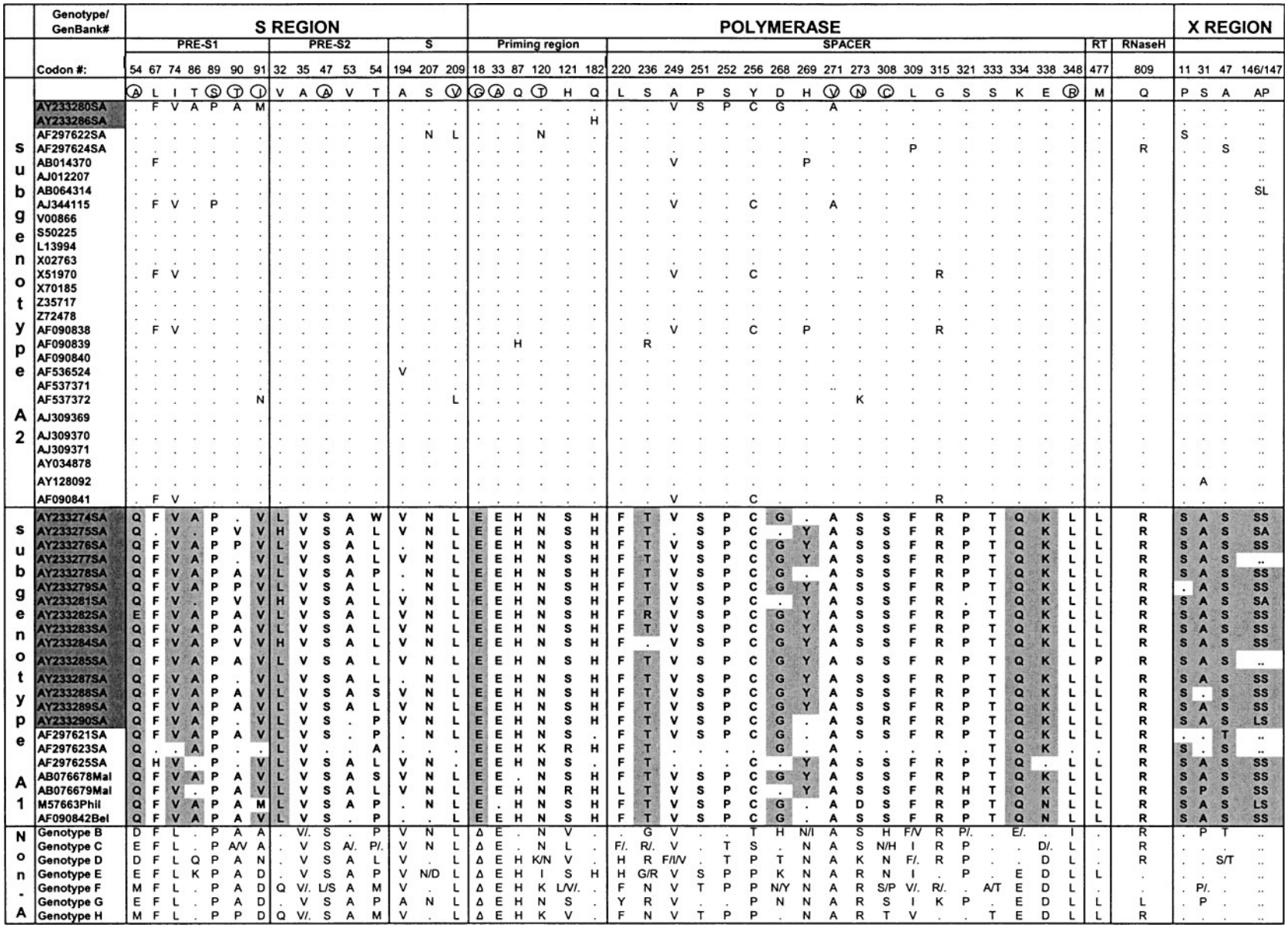

Fig. 2. For legend see page 1218.

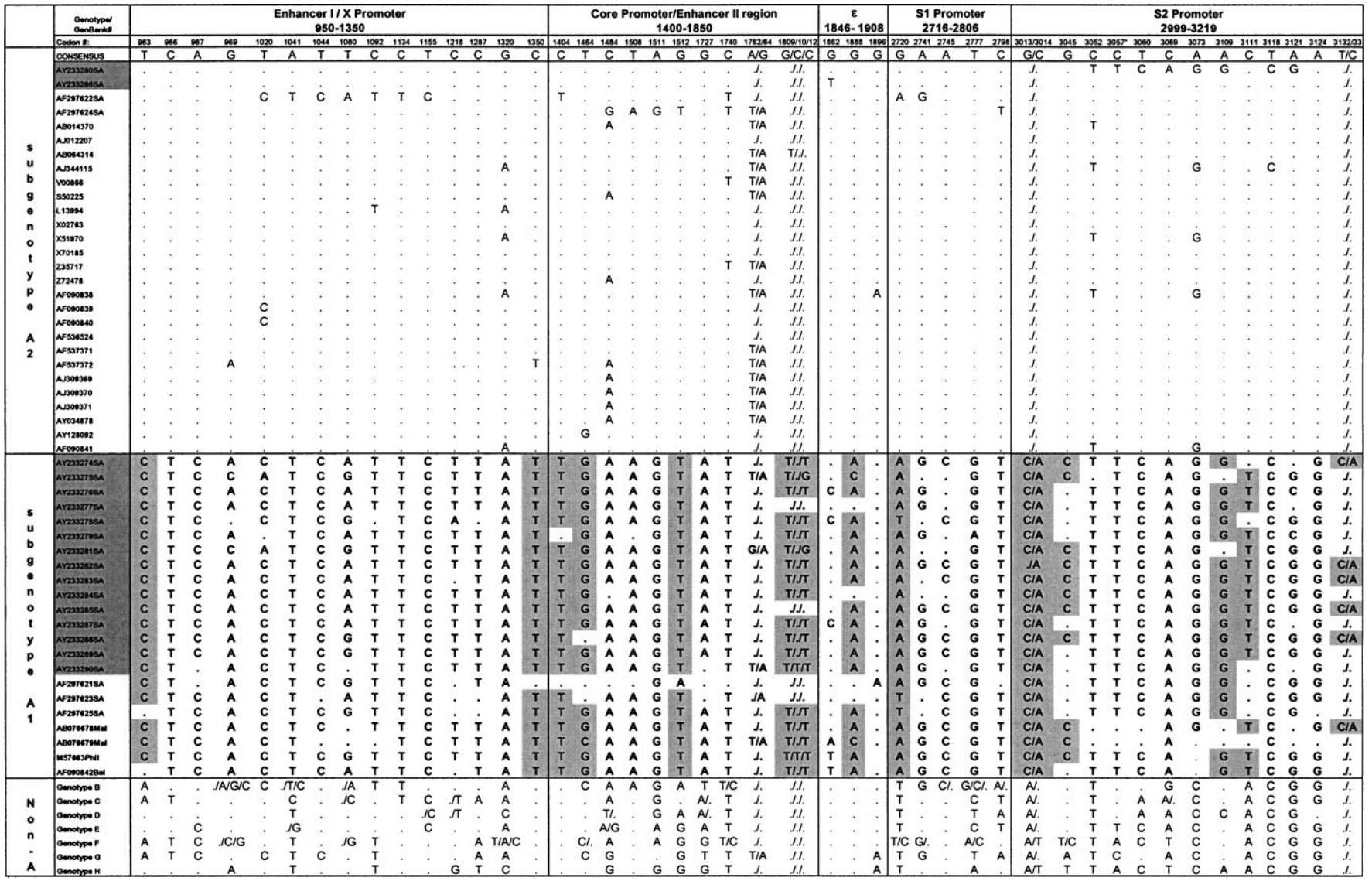

Fig. 3. For legend see page 1218.

Table 3. Mutations within the *cis*-acting regulatory elements found predominantly in subgenotype A1

Numbering of nucleic acids according to genotype A [AY233276] and functional domains according to Moolla *et al.* (2002).

cis-acting element	Mutation	Functional element affected
Enhancer I/X promoter	A/T963C	5′ modulator element
Core promoter	C1404T	NRE-γ
	C/T1464G	NRE-β
	A/G1512T	NRE-β
	G1809T, A1811T, C1812T/G	precore Kozak Sequence*
ε	G1888A	encapsidation signal
S1 promoter	T/G2720A	HNF-1 binding site
S2 promoter	[A/G3013C,C/T3014A]	NF1 transcription factor-binding site
	T/A/G3045C	region A
	C/A3109G, A/C3111T	region E
	[T3132C,C3133A]	region F

*Shown to reduce HBeAg translation by ribosomal leaky scanning mechanism (Ahn *et al.*, 2003).

the protein coding regions can also lead to alterations in these elements. Table 3 shows the mutations within the regulatory elements that are characteristic of subgenotype A1 and the regions of the *cis*-acting elements that are affected by the mutations. Although it is not possible to deduce the effect of these changes on the replication of the virus, it is probable that they have a role to play in modulating replication and resulting in the reduced HBV DNA levels that are found in patients infected with subgenotype A1 (Kramvis *et al.*, 1997).

The silent G to A nucleotide mutation at position 1888 was unique to subgenotype A1 (Fig. 3). This mutation occurs rarely in other genotypes and in HBV isolates from outside Africa. In addition to stabilizing the encapsidation signal (ε) (Kramvis & Kew, 1998) and possibly affecting reverse transcription, this mutation could affect the translation of the core protein. The 1888 G to A mutation introduces an out-of-frame AUG start codon, 13 nucleotides upstream of the core AUG start codon and a minicistron that can potentially be translated into seven amino acids: 'Met-Ala-Leu-Gly-His-Gly-His'. Therefore, the newly introduced start codon at 1888 may have an important role in the regulation of the translational efficiency of a downstream start codon (Rogozin *et al.*, 2001), in our case the start codon for translation of the core protein. The presence of small upstream ORFs in the leader sequence has also been found to have a modulating role in the translation of proteins from downstream cistrons in a number of viruses (Biegalke & Geballe, 1990; de Smit & van Duin, 1993; Degnin *et al.*, 1993; Ozawa *et al.*, 1988; Ryabova *et al.*, 2002). Similarly, the translation of HBV polymerase gene is

Fig. 2. (*on page 1216*) Comparison of amino acid residues of S, polymerase and X ORFs of subgenotype A1 isolates with amino acid sequences of subgenotype A2 and other HBV genotypes. *Dots* indicate amino acid identity and Δ an amino acid deletion. Amino acid residue 1 in the pre-S1 refers to the first amino acid of genotype A sequences. SA isolates sequenced in the present study are shaded in *grey* and all other sequences were obtained from GenBank. For the non-A genotypes the representative amino acid was determined by aligning 10 sequences of each genotype from geographically distinct regions and the consensus amino acid deduced if it occurred in 60% or more of the sequences. Amino acids found only in subgenotype A1 and in other non-A genotypes (B-H) but not subgenotype A2, are shown in **bold** and those found predominantly in subgenotype A1 and not in subgenotype A2 nor in other genotypes are in bold and are shaded in *grey*. Amino acids unique to subgenotype A2 are *circled*.

Fig. 3. (*on page 1217*) Comparison of the nucleic acid sequences of the *cis*-acting elements of subgenotype A1 isolates with sequences of subgenotype A2 and other HBV genotypes. *Dots* indicate amino acid identity. Nucleotide 1 denotes the nucleotide position of hepatitis B virus *adw* genome (GenBank accession # AY233276) where the *Eco*RI cleavage site is position 1. SA isolates sequenced in the present study are shaded in *grey* and all other sequences were obtained from GenBank. For the non-A genotypes the representative nucleic acid was determined by aligning 10 sequences of each genotype from geographically distinct regions and the consensus nucleic acid deduced if it occurred in 60% or more of the sequences. Nucleotides found only in subgenotype A1 and in other non-A genotypes (B-H) but not subgenotype A2, are shown in **bold** and those found predominantly in subgenotype A1 and not in subgenotype A2 nor other genotypes are in bold and are shaded in *grey*.

controlled by a leaky scanning mechanism together with a termination-reinitiation mechanism involving an upstream minicistron (Fouillot *et al.*, 1993; Hwang & Su, 1998). Therefore, it is conceivable that the introduction of the start codon by the 1888 G to A mutation, seen in subgenotype A1, may play a modulating role in the translation of the core protein and needs further investigation.

In conclusion, it can be seen that subgenotype A1 HBV isolates from SA differ from subgenotype A2 in two ways. Firstly, subgenotype A1 isolates have distinctive sequence characteristics that may affect both the replication of the virus and the expression of its proteins. Secondly, the mean nucleotide divergence of subgenotype A1 is greater than that for subgenotype A2 suggesting that subgenotype A1 has been endemic and has a very long natural history within the South African black population.

ACKNOWLEDGEMENTS

Anna Kramvis and Gerald Kimbi received grants from the Poliomyelitis Research Foundation of South Africa and the H. E. Griffin Cancer Trust.

REFERENCES

Ahn, S., Kramvis, A., Kawai, S., Spangenberg, H., Li, J., Kimbi, G., Kew, M., Wands, J. & Tong, S. (2003). Sequence variation upstream of the precore AUG start codon of hepatitis B virus reduces HBe-antigen production. *Gastroenterology* **125**, 1370–1378.

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

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.

Arii, M., Takada, S. & Koike, K. (1992). Identification of three essential regions of hepatitis B virus X protein for trans-activation function. *Oncogene* **7**, 397–403.

Baptista, M., Kramvis, A. & Kew, M. C. (1999). High prevalence of 1762^{T} 1764^{A} mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. *Hepatology* **29**, 946–953.

Biegalke, B. J. & Geballe, A. P. (1990). Translational inhibition by cytomegalovirus transcript leaders. *Virology* **177**, 657–667.

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. *J Gen Virol* **78**, 1719–1729.

Das, K., Xiong, X., Yang, H., Westland, C. E., Gibbs, C. S., Sarafianos, S. G. & Arnold, E. (2001). Molecular modeling and biochemical characterization reveal the mechanism of hepatitis B virus polymerase resistance to lamivudine (3TC) and emtricitabine (FTC). *J Virol* **75**, 4771–4779.

Degnin, C. R., Schleiss, M. R., Cao, J. & Geballe, A. P. (1993). Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol* **67**, 5514–5521.

de Smit, M. H. & van Duin, J. (1993). Translational initiation at the coat-protein gene of phage MS2: native upstream RNA relieves inhibition by local secondary structure. *Mol Microbiol* **9**, 1079–1088.

Estacio, R. C., Chavez, C. C., Okamoto, H., Lingao, A. L., Reyes, M. T., Domingo, E. & Mayumi, M. (1988). Nucleotide sequence of a hepatitis B virus genome of subtype *adw* isolated from a Philippino: comparison with the reported three genomes of the same subtype. *J Gastroenterol Hepatol* **3**, 215–222.

Felsenstein, J. (1995). PHYLIP (Phylogeny Inference Package). Department of Genetics, University of Washington, Seattle, Washington. http://evolution.genetics.washington.edu/phylip/software.html.

Fouillot, N., Tlouzeau, S., Rossignol, J. M. & Jean-Jean, O. (1993). Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation. *J Virol* **67**, 4886–4895.

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

Gunther, S., Li, B. C., Miska, S., Kruger, 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.

Hwang, W. L. & Su, T. S. (1998). Translational regulation of hepatitis B virus polymerase gene by termination-reinitiation of an upstream minicistron in a length-dependent manner. *J Gen Virol* **79**, 2181–2189.

Kidd-Ljunggren, K., Oberg, M. & Kidd, A. H. (1995). The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. *J Gen Virol* **76**, 2119–2130.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* **196**, 947–950.

Kramvis, A. & Kew, M. C. (1998). Structure and function of the encapsidation signal of hepadnaviridae. *J Viral Hepat* **5**, 357–367.

Kramvis, A. & Kew, M. C. (1999). The core promoter of hepatitis B virus. *J Viral Hepat* **6**, 415–427.

Kramvis, A., Bukofzer, S., Kew, M. C. & Song, E. (1997). Nucleic acid sequence analysis of the precore region of hepatitis B virus from sera of southern African black adult carriers of the virus. *Hepatology* **25**, 235–240.

Kramvis, A., Weitzmann, L., Owiredu, W. K. & 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.

Laskus, T., Rakela, J., Steers, J. L., Wiesner, R. H. & Persing, D. H. (1994). Precore and contiguous regions of hepatitis B virus in liver transplantation for end-stage hepatitis B. *Gastroenterology* **107**, 1774–1780.

Mayerat, C., Mantegani, A. & Frei, P. C. (1999). Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection? *J Viral Hepat* **6**, 299–304.

Moolla, N., Kew, M. & Arbuthnot, P. (2002). Regulatory elements of hepatitis B virus transcription. *J Viral Hepat* **9**, 323–331.

Mphahlele, M. J., Francois, G., Kew, M., Van Damme, P., Hoosen, A. & Meheus, A. (2002). Epidemiology and control of hepatitis B: implications for eastern and southern Africa. *S Afr J Epidemiol Infect* **17**, 12–17.

Nassal, M. & Schaller, H. (1996). Hepatitis B virus replication – an update. *J Viral Hepat* **3**, 217–226.

Neurath, A. R., Kent, S. B., Strick, N. & Parker, K. (1986). Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. *Cell* **46**, 429–436.

Nicholas, K. B. & Nicholas, H. B., Jr (1997). GeneDoc: analysis and visualization of genetic variation. http://www.psc.edu/biomed/genedoc.

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

Norder, H., Hammas, B., Lofdahl, S., Courouce, 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. *J Gen Virol* **73**, 1201–1208.

Norder, H., Hammas, B., Lee, S. D., Bile, K., Courouce, 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. *J Gen Virol* **74**, 1341–1348.

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* **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.

Odemuyiwa, S. O., Mulders, M. N., Oyedele, O. I., Ola, S. O., Odaibo, G. N., Olaleye, D. O. & Muller, C. P. (2001). Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa. *J Med Virol* **65**, 463–469.

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.

Owiredu, W. K., Kramvis, A. & Kew, M. C. (2001). Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B surface antibody alone: possible association with recombination between genotypes A and D. *J Med Virol* **64**, 441–454.

Ozawa, K., Ayub, J. & Young, N. (1988). Translational regulation of B19 parvovirus capsid protein production by multiple upstream AUG triplets. *J Biol Chem* **263**, 10922–10926.

Page, R. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358. http://taxonomy.zoology.gla.ac.uk/rod/treeview.html.

Pontisso, P., Ruvoletto, M. G., Gerlich, W. H., Heermann, K. H., Bardini, R. & Alberti, A. (1989). Identification of an attachment site for human liver plasma membranes on hepatitis B virus particles. *Virology* **173**, 522–530.

Rogozin, I. B., Kochetov, A. V., Kondrashov, F. A., Koonin, E. V. & Milanesi, L. (2001). Presence of ATG triplets in 5′ untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics* **17**, 890–900.

Ryabova, L. A., Pooggin, M. M. & Hohn, T. (2002). Viral strategies of translation initiation: ribosomal shunt and reinitiation. *Prog Nucleic Acid Res Mol Biol* **72**, 1–39.

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., Kato, H. & 8 other authors (2003). Genotype, serotype, and phylogenetic characterization of the complete genome sequence of hepatitis B virus isolates from Malawian chronic carriers of the virus. *J Med Virol* **69**, 33–40.

Takahashi, K., Akahane, Y., Hino, K., Ohta, Y. & Mishiro, S. (1998). Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates. *Arch Virol* **143**, 2313–2326.

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

Uy, A., Wunderlich, G., Olsen, D. B., Heermann, K. H., Gerlich, W. H. & Thomssen, R. (1992). Genomic variability in the preS1 region and determination of routes of transmission of hepatitis B virus. *J Gen Virol* **73**, 3005–3009.

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. *J Gen Virol* **82**, 893–897.

Vieth, S., Manegold, C., Drosten, C., Nippraschk, T. & Gunther, S. (2002). Sequence and phylogenetic analysis of hepatitis B virus genotype G isolated in Germany. *Virus Genes* **24**, 153–156.

WHO (2002). Hepatitis B. World Health Organization Department of Communicable Diseases Surveillance Reponse. WHO/CDS/CSR/LYO/2002.2: Hepatitis B. http://www.who.int/emc-documents/hepatitis/docs/whocdscsrlyo20022.pdf.

Xia, X. (2000). *Data Analysis in Molecular Biology and Evolution.* Dordrecht: Kluwer.