Journal of General Virology (1997), **78**, 1719–1729. Printed in Great Britain

A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa

Sheila M. Bowyer,[1] Louise van Staden,[2] Michael C. Kew[2] and John G. M. Sim[1]

[1] National Institute for Virology and Department of Medical Virology, University of the Witwatersrand, Private Bag X4, Sandringham 2131, South Africa

[2] MRC Molecular Hepatology Research Unit, Department of Medicine, University of the Witwatersrand Medical School, 7 York Road, Parktown, Gauteng 2193, South Africa

The preS2/S genes of hepatitis B virus isolated from 29 acutely or chronically infected individuals in the Gauteng province of South Africa were sequenced. Phylogenetic analysis of these sequences in comparison with global isolates from the GenBank database showed that 24 sequences clustered with genotypic group A, three with genotypic group D and one each with genotypic groups B and C. Group A isolates had greater identity with groups D (variation of 6·6%) and E (6·8%) than with the Eastern groups B (7·4%) and C (8·1%) and were most different from group F (11·0%). Of the South African group A specimens, 59·1% clustered with two global sequences to form a discrete segment which we have called subgroup A′. The amino acid differences that set these isolates apart from the rest of group A tended to cluster in the preS2 region (amino acids 7, 10, 32, 35, 47, 48, 53 and 54), with a few changes occurring in the major surface antigen (amino acid sites 207 and 209). Analysis of isolates showed that there was a 9-fold higher prevalence of the *ay* determinant in South Africa than previously reported.

Introduction

Hepatitis B virus (HBV) strains are divided into nine major serotypes, *adw2*, *adw4* *q*−, *ayw1*, *ayw2*, *ayw3*, *ayw4*, *adr* *q*+, *adr* *q*− and *ayr* (Couroucé *et al.*, 1976; Couroucé-Pauty *et al.*, 1978). These divisions are based on a common *a* determinant and two mutually exclusive *d*/*y* and *r*/*w* determinants of the major envelope protein, and have been widely used to differentiate HBV strains in clinical, virological and epidemiological studies (Norder *et al.*, 1993). Serotypes have been shown to be geographically distributed (Couroucé-Pauty & Soulier, 1983). For example, *adr* is confined to the East, *ayw2* predominates in the Mediterranean region and *adw4* is found only in the Pacific region. Serotype *adw2* predominates in Northern Europe and sub-Saharan Africa, but its prevalence declines moving from East Africa towards Central and West Africa and is accompanied by a corresponding increase in the unique African serotype, *ayw4* (Magnius & Norder, 1995). Of HBV serotypes in South Africa, 97·7% were previously reported to be serotype *adw2*, with the remaining 2·3% being *ayw* (Couroucé-Pauty & Soulier, 1983). Serotypic differences are based on discrete and limited sequence changes and thus it is not surprising that when isolates of known serotype are sequenced, no genetic relatedness is evident (Yang *et al.*, 1995). Alignment, comparison and phylogenetic analysis of sequence data from various parts of the world show that HBV clusters into six genotypic groups, A to F (Norder *et al.*, 1993). Several of the genotypic groups are serotypically heterogeneous: group C contains *adw*, *adr* and *ayr*; both groups A and B contain serotype *adw* and *ayw1*; and group D contains serotypes *ayw2*, *ayw3* and *ayw4*. In addition, the genotypic groups are also themselves geographically arranged (Magnius & Norder, 1995): *adw2* clusters into group A in the West but into groups B and C in the East; *ayw1* also occurs in group A in the West, but is confined to group B in the East; group A is seldom found in the East, whereas group C is limited to the East; group E is limited to parts of East, Central and West Africa and group F to the Pacific (Magnius & Norder, 1995). Little is known about the HBV genotypes that occur in southern Africa. To date the only published sequences have been those integrated into the PLC/PRF/5 hepatocellular carcinoma cell line (Rivkina *et al.*, 1988) and the complete preS/S gene of a clone from Zimbabwe (Chirara & Chetsanga, 1994). We have therefore sequenced the preS2 and S gene regions of HBV isolates from patients with acute hepatitis B and from chronic carriers of the virus in the Gauteng province of South Africa and compared our data with published sequences.

Author for correspondence: Michael C. Kew.
Fax +27 11 643 8777. e-mail 014KED@CHIRON.WITS.AC.ZA

Table 1. Designation of 110 HBV isolates used in this study

No.	Patient/strain	Accession no./reference	Genotype	Serotype	Origin
A01	HHVBAMAM	X75669	A	ayw1*	Cameroon
A02	HPBS2ADW	M12346	A	adw*	The East
A03	S50225	S50225	A	adw	UK
A04	HPBSSWT	M74498	A	adw2	France
A05	HPBSSB1MUT	M74499	A	adw2	France
A06	HHVBA	X75666	A	adw2*	France
A07	HVHEPB	X51970	A	adw*	Germany
A08	HBVXCPS	X70185	A	adw2*	Germany
A09	HPBS	M21030	A	adw2*	Mozambique
A10	LS	Kidd-Ljünggren†	A	adw2*	The Philippines
A11	JOA	Kidd-Ljünggren†	A	adw2*	Poland
A12	MP	Kidd-Ljünggren†	A	adw2*	Poland
A13	HUMPRECX	L13994	A	adw*	Eastern Europe
A14	1385	van Staden (U87739)‡	A	adw*	South Africa
A15	A24	Bowyer (U87725)‡	A	adw	South Africa
A16	P08	Bowyer (U87726)‡	A	adw	South Africa
A17	E14	Bowyer (U87727)‡	A	adw	South Africa
A18	H510	van Staden (U87740)‡	A	adw*	South Africa
A19	E20	Bowyer (U87728)‡	A	adw*	South Africa
A20	7983	van Staden (U87742)‡	A	adw*	South Africa
A21	P03	Bowyer (U87729)‡	A	adw	South Africa
A22	O04	Bowyer (U87730)‡	A	adw*	South Africa
A23	E17	Bowyer (U87731)‡	A	adw*	South Africa
A24	O01	Bowyer (U87732)‡	A	adw	South Africa
A25	E30	Bowyer (U87733)‡	A	adw*	South Africa
A26	1019	van Staden (U87741)‡	A	adw*	South Africa
A27	1149	van Staden (U87852)‡	A	adw*	South Africa
A28	8017	van Staden (U87744)‡	A	adw*	South Africa
A29	1449	van Staden (U87745)‡	A	adw	South Africa
A30	IR	van Staden (U87743)‡	A	adw*	South Africa
A31	O02	Bowyer (U88094)‡	A	adw*	South Africa
A32	E11	Bowyer (U87734)‡	A	adw	South Africa
A34	E34	Bowyer (U88095)‡	A	adw*	South Africa
A36	A23	Bowyer (U87735)‡	A	adw	South Africa
A38	S01	Bowyer (U88096)‡	A	adw	South Africa
A39	E06	Bowyer (U87736)‡	A	ayw*	South Africa
A40	7782	van Staden (U87746)‡	A	adyw*	South Africa
A41	HPBVSAG2	M54898	A	adw*	Taiwan
A42	HBADWZCG	M57663	A	adw2*	The Philippines
A43	HBVADW	V00866	A	adw2*	USA
A44	HBVADW2	X02763	A	adw2*	USA
A45	HBVPRESS	X69458	A	ayw*	Zimbabwe
A46	HPBSSB1REC	M74500	A	adw2	France
B01	HPBADW3	D00331	B	adw3*	Indonesia
B02	HPBADW1	D00329	B	adw2*	Japan
B03	HPBADW2	M54923	B	adw2*	Indonesia
B04	HPBADW2	D00330	B	adw2*	Japan
B05	WONG	van Staden (U87747)	B	adw*	South Africa
B06	ISWARI	Lauder§	B	adw	
B07	HVBBS	X75660	B	ayw1*	
C01	OKAADW	Lauder§	C	adw*	
C02	HPBADR1CG	M38454	C	adr*	China
C03	CP	Kidd-Ljünggren†	C	adr*	China
C04	S62754	S62754	C	adw*	The East
C05	HPBSAYRA	M17688	C	ayr*	The East
C06	HHVBCS	X75792	C	adr*	France
C07	DD	Kidd-Ljünggren†	C	adr*	Indonesia
C08	HPBCG	D12980	C	adr*	Japan
C09	HBVADR	V00867	C	adr*	Japan
C10	HPBETNC	L08805	C	adr*	Japan
C11	HPBADRA	M12906	C	adr*	Japan
C12	HBVADR4	X01587	C	adr*	Japan

Table 1. (*cont.*)

No.	Patient/strain	Accession no./reference	Genotype	Serotype	Origin
C13	HPBHBSAGA	M23805	C	adr*	Japan
C14	HPBHBSAGB	M23806	C	adr*	Japan
C15	HPBHBSAGD	M23808	C	adr*	Japan
C16	HEHBVAYR	X04615	C	ayr*	The East
C17	HPBCGADR	M38636	C	adr*	Korea
C18	HBVADRM	X14193	C	adr*	Korea
C19	KRJ	Kidd-Ljünggren†	C	adr*	Korea
C20	X75665	X75665	C	adrq−*	New Caledonia
C21	491	Kidd-Ljünggren†	C	adr*	New Zealand
C22	651	Kidd-Ljünggren†	C	adr*	New Zealand
C23	029	Kidd-Ljünggren†	C	adr*	New Zealand
C24	HHVCCHA	X75656	C	adrq−*	Polynesia
C25	LEE	van Staden (U87748)	C	adr*	South Africa
C26	AP	Kidd-Ljünggren†	C	adr*	Thailand
D01	HPBAYW	J02203	D	ayw3*	France
D02	HHVBD	X75668	D	ayw3*	France
D03	HHVBDS	X75662	D	ayw2*	France
D04	HBVORFS	X72702	D	ayw*	Germany
D05	PC	Kidd-Ljünggren†	D	ayw*	India
D06	NB	Kidd-Ljünggren†	D	ayw*	Iran
D07	HBVAYWCI	X65258	D	ayw*	Italy
D08	HBVAYWE	X65259	D	ayw*	Italy
D09	HBVAYWC	X65257	D	ayw*	Italy
D10	HBVAYWMCG	X59795	D	ayw*	Italy
D11	HBVDNA	X68292	D	ayw*	Italy
D12	TY5	Kidd-Ljünggren†	D	ayw2*	Lebanon
D13	XXHEPAV	X02496	D	ayw*	Northern Europe
D14	376	Kidd-Ljünggren†	D	ayw*	New Zealand
D15	567	Kidd-Ljünggren†	D	ayw2*	New Zealand
D16	FEMI	Kidd-Ljünggren†	D	ayw*	Romania
D17	HPBHBSAG	M12393	D	ayw*	Eastern Europe
D18	HPBADYW	J02202	D	adyw*	UK
D19	MAJ	Kidd-Ljünggren†	D	ayw2*	Somalia
D20	E08	Bowyer (U87737)	D	ayw*	South Africa
D21	S09	Bowyer (U87738)	D	ayw*	South Africa
D22	P67	Bowyer (U87851)	D	ayw*	South Africa
D25	HPBHBVAA	M32138	D	ayw2*	Turkey
D26	AN	Kidd-Ljünggren†	D	ayw*	Turkey
E01	YA	Kidd-Ljünggren†	E	ayw4*	Liberia
E02	JAM	Kidd-Ljünggren†	E	ayw4*	Liberia
E03	HPBVAR	L24071	E	ayw*	Northern Europe
E04	HHVBE4	X75664	E	ayw4*	Senegal
E05	HPBVCG	D00220	E	adw2*	West Africa
E06	HHVBBAS	X75657	E	ayw4*	West Africa
F01	HBVADW4A	X69798	F	adw4*	Brazil
F02	HHVBF	X75663	F	adw4q−*	Colombia
F03	HHVBFFOU	X75658	F	adw4q−*	France
F04	HHVBFS	X75661	F	adw4q−*	France

* South African isolates that were serologically subtyped or specimens whose serotype was specifically stated in the reference; all other serotypes were deduced from the sequence.
† Sequences from Kidd-Ljünggren *et al.* (1994).
‡ Sequences from this paper.
§ Sequences from Lauder *et al.* (1993).
‖ Country of origin of the research institution where the nationality of the patient is not specifically cited.

Methods

■ **Patients.** HBV DNA was extracted from the serum of 29 South African patients, 11 with acute hepatitis B and 18 chronic carriers of the virus. Of the patients studied, 24 were male and five were female; 17 were Black, four of mixed descent (Eurafrican), five Caucasian, one Indian and two Chinese. The complete sequence of the preS2/S gene was obtained from 26 of the patients; two isolates were sequenced in the preS2 region only, and one in the S gene region only.

■ **HBV DNA extraction.** HBV DNA was extracted from 200 μl serum

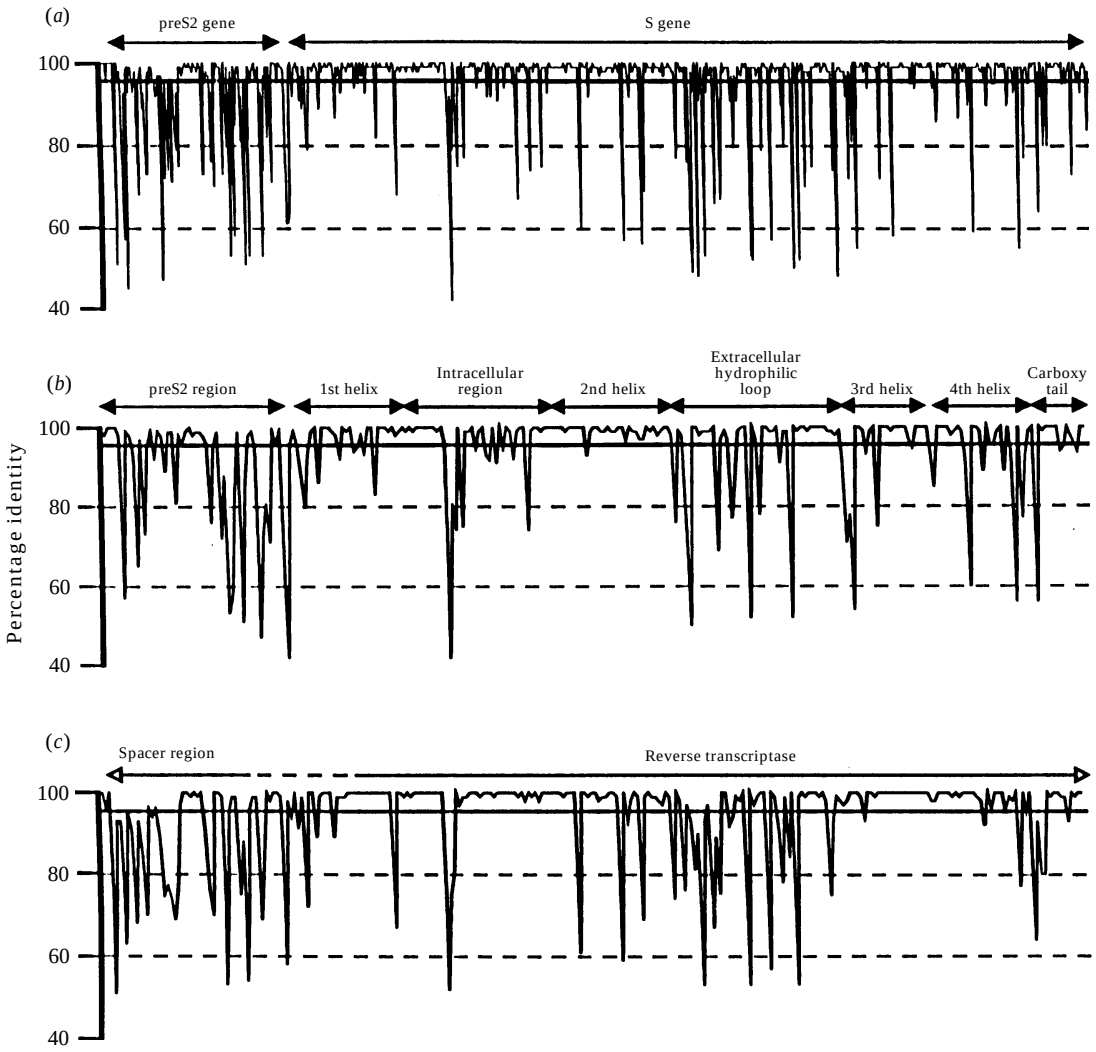

Fig. 1. Percentage occurrence of the most common entity plotted at each position in the nucleotide/amino acid preS2/S consensus sequences. (*a*) Nucleotides, (*b*) amino acids coding for the middle and small surface proteins (preS2/S sequence) and (*c*) amino acids coding for the polymerase gene. The horizontal line indicates 95% identity.

using QIAamp rapid purification columns (Qiagen) according to the manufacturer's instructions.

■ **DNA amplification.** For the acute hepatitis B patients, the complete preS/S gene was amplified using sense primer 2810 (numbering from the *Eco*RI site) (5′ CACGTAGCGCCTCATTTTCTGGGTCACCATATTCT 3′) and antisense primer 979 (5′ CAAAAGACCCACAATTCTTTGACATACTTTCCAAT 3′) to yield a 1403 bp amplicon. The proofreading enzyme *Pfu* (0·0025 U/μl) was added to the PCR reaction mix [10 mM Tris–HCl pH 9·0, 50 mM KCl, 1·5 mM $MgCl_2$, 200 μM of each dNTP, 0·05 U/μl *Taq* polymerase (Promega), 1 μM of each primer, 0·1% Triton-X100] to enhance the fidelity of the polymerase reaction. The entire core and surface genes of the chronic carriers were amplified using sense primer 1738 (5′ AGTTGGGGGAGGAGATTAG 3′) and antisense primer 1018 (5′ CCACATTGTGTAAATGGAGCAGC 3′) to yield a 2483 to 2522 bp fragment (depending on genotype) using *TaqExpand* polymerase enzyme (Boehringer), which contains proofreading activity. When no product was evident on first round PCR, a nested PCR was performed using sense primer 2833 (5′ CTTGGGAACAAGAGCTACAGCAT 3′) and antisense primer 1018, as above. This gave a 1388 to 1427 bp amplicon.

■ **DNA sequencing.** Oligonucleotides and excess dNTPs were removed from the PCR products either with Magic PCR Preps (Promega) or by digestion with Exonuclease I and alkaline phosphatase according to the protocol from the USB direct sequencing kit (Amersham). The preS2 (from nt −11 to 154) and the S genes (from nt 155 to 845) (coding for the middle and major surface antigens) of all amplicons were sequenced directly (USB direct sequencing kit, Amersham) in both forward and reverse directions.

■ **Serotyping.** Serotypes for all samples were deduced from sequence data and some of these were confirmed using monoclonal antibodies directed against the *a*, *d*, *y*, *w* and *r* determinants of the surface antigen (serotyped South African specimens are marked with an asterisk in Table 1) using the HBsAg subtype kit from the Institute of Immunology Co., Tokyo, Japan.

■ **Data analysis.** Twenty-six South African isolates were compared with 42 sequences from GenBank, complete over the relevant 846 nucleotides. The remaining three South African isolates, which lacked some sequence information at both ends of the fragment, were analysed together with 23 'preS2 only' or 16 'major surface gene only' sequences from GenBank. These sequences, together with our 26 full sequences and the 42 GenBank sequences, enabled a comparison to be made of 93 specimens in the preS2 region and 85 in the S gene region. All sequences used in this study are referenced in Table 1. Mutational hot spots in the sequenced region were ascertained by storing all sequence information in a database (dBase III+, Ashton Tate) and then calculating the frequency of each base at each nucleotide position using a customized program. Similarly, the frequency at each site of each amino acid in the deduced proteins, surface and polymerase, was calculated. The nucleotide and amino acid consensus sequences (the latter in both reading frames) were obtained from these frequency tables, and the percentage occurrence of the most common entity was plotted for each position in the genome (Fig. 1). Dendrograms were generated using PHYLIP (Phylogeny Inference Package) version 3.5c DNAPARS, FITCH, DNAML, DRAWTREE and DRAWGRAM software programs (Felsenstein, 1993).

Results

Fig. 1 illustrates the identity, at each site of the preS2/S gene at the nucleotide and amino acid level, of both transcribed open reading frames (ORF) (i.e. the middle surface protein and part of the polymerase enzyme). The overall nucleotide and amino acid variations for all sequences compared in this study appear as troughs in this graph, the depth of which reflects the amount of variation at a particular site. Variable amino acid sites were defined as positions at which the predominant amino acid was present at a frequency of less than 95%. Using these criteria all of the previously reported determinants were observed. Using these data, the degree of diversity of important preS2 and S domains was calculated, both at the nucleotide and amino acid level (Table 2).

The 68 complete preS2/S sequences were analysed using DNAPARS and DRAWGRAM to produce a horizontal phylogram (results not shown). The same sequences were then used as input for DNADIST and the output distance matrix fed into FITCH and DRAWTREE (Fig. 2) to produce a tree showing phylogenetic distances between the sequences. The South African samples cluster predominantly with the known genotypic groups A and D. Of the five serotype *ayw* South African isolates, three cluster into genotypic group D and two into group A. The rest of group A are serotype *adw*. Sequence data from two patients, both of Chinese origin, do not cluster with the rest of the specimens but cluster into group B (*adw*) and group C (*adr*), respectively.

To improve the accuracy of the trees, the two genotypic groups of interest, namely group A and group D, were analysed individually using the non-distance matrix program DNAML (Fig. 3). Both Fig. 2 and Fig. 3(*a*) show clearly that group A splits into two distinct subgroups. Of the South African isolates, 40·9% (A14, A15, A16, A21, A22, A24, A26, A28 and A36) cluster with published sequences, forming the first subgroup of A, while the remaining 59·1% (A17, A18, A19, A20, A23, A25, A27, A29, A30, A32, A39 and A40) cluster separately to form a discrete genotypic segment, which we referred to as subgroup A′. Also clustering in subgroup A′ were three isolates from GenBank, A45 (a Zimbabwean isolate), A04 (a French isolate; Tran *et al.*, 1991) and A42 (a clone isolated from the Philippines; Estacio *et al.*, 1988). A05 and A46, which are variants of A04, also cluster with subgroup A′.

Pairwise analysis of nucleotide divergence in the 32 complete preS2/S group A sequences of HBV DNA was performed and the mean difference between these two segments of group A was found to be 3·8% whilst the difference within each subgroup was 1·7% (group A) and 2·9% (group A′). Similarly, intra- and intergroup nucleotide and amino acid divergence of all the groups was calculated (Fig. 4). From this figure, it can be seen that group A has greater

Table 2. Percentage diversity at the nucleotide and amino acid level across the different domains spanning the preS2/S gene open reading frame

nt no.	nt (% diversity)	PreS2/S aa spanned	S (% diversity)	Domains of preS2 and S
3211–154	36·3%	1–55	43·6%	preS2 gene
155–175	25%	1–7	42·8%	Amino terminal of S
176–238	11%	8–28	29%	First membrane-spanning helix of S
239–382	15·3%	29–76	23%	Intracellular loop of S
383–448	6%	77–98	4·5%	Second membrane-spanning helix of S
449–631	20·3%	99–159	20%	Hydrophilic extracellular loop
632–706	13·3%	160–184	28%	Third membrane-spanning helix of S
707–718	0%	185–188	0%	Short region between helices
719–781	19%	189–209	33·3%	Fourth membrane-spanning helix of S
782–835	22·2%	210–227	29%	Carboxy tail

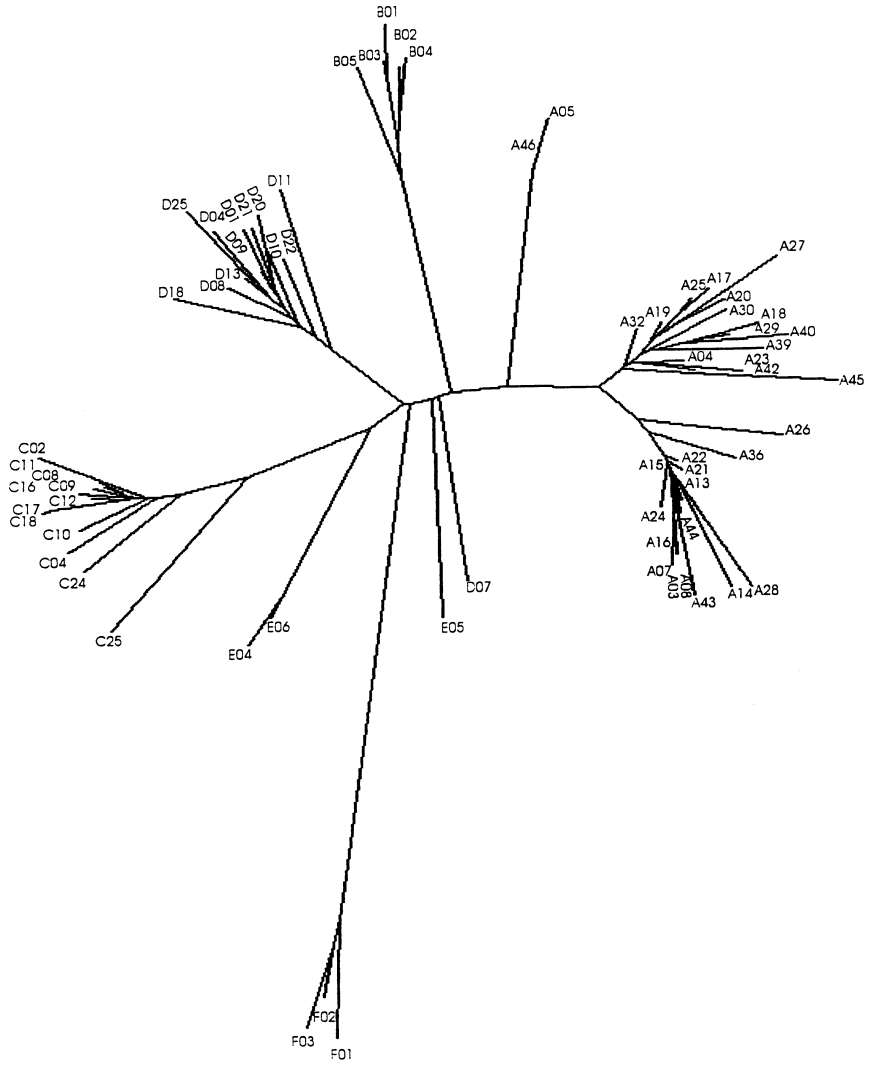

Fig. 2. Phylogenetic distance tree produced using the PHYLIP programs FITCH and DRAWTREE from preS2/S sequence data from 26 South African sequences and 42 global sequences from GenBank.

identity with groups D (6·6%) and E (6·8%) than with the Eastern groups B (7·4%) and C (8·1%), and is most different from group F (11·0%).

Discussion

Using the preS2/S gene nucleotides, which constitute 26·3% of the total genome, we have genotyped 26 HBV isolates from acutely or chronically infected individuals. An additional three PCR amplicons lacking either the preS2 region (one specimen) or part of the S gene (two specimens) were also successfully genotyped. The degree of diversity (Table 2) of the preS2 region at both the nucleotide and amino acid levels is greater than that of the S region. When the preS2 region alone (constituting 5·1% of the genome) was used for the analysis, this short stretch of 165 nucleotides was found to be sufficient for the definition of the six HBV genotypic groups.

Amino acid alignment revealed sites where characteristic clustering of variants occurred (Fig. 5). Some of these sites have not been reported previously. In particular, a series of variations in subgroup A′ sequences, predominantly in the preS2 region, set these isolates apart from the rest of group A, which is otherwise well conserved.

Cysteines at sites vital for the maintenance of the antigenicity of the major surface antigen (codons 48, 65, 69, 121, 124, 137, 138, 139, 147 and 149; Antoni *et al.*, 1994) were well conserved in our samples and no changes at previously reported vaccine escape-mutant sites were found in the database (Met[133], Oon *et al.*, 1995; Lys[141], Whittle, *et al.*, 1991; Pro[142], Ashton-Rickardt & Murray, 1989; Ala[144], Harrison, *et al.*, 1991; Gly[145], Waters *et al.*, 1992 and Fujii *et al.*, 1992).

In the past, specimens were categorized by serotype but since serotype specificities are defined by point mutations, it is far more informative to sequence longer stretches of the genome and then assign the isolate to a genotype by phylogenetic analysis. Both the 165 nucleotides of the preS2 region or the 681 nucleotides of the S gene (or a combination of both) were found to be suitable for genotyping.

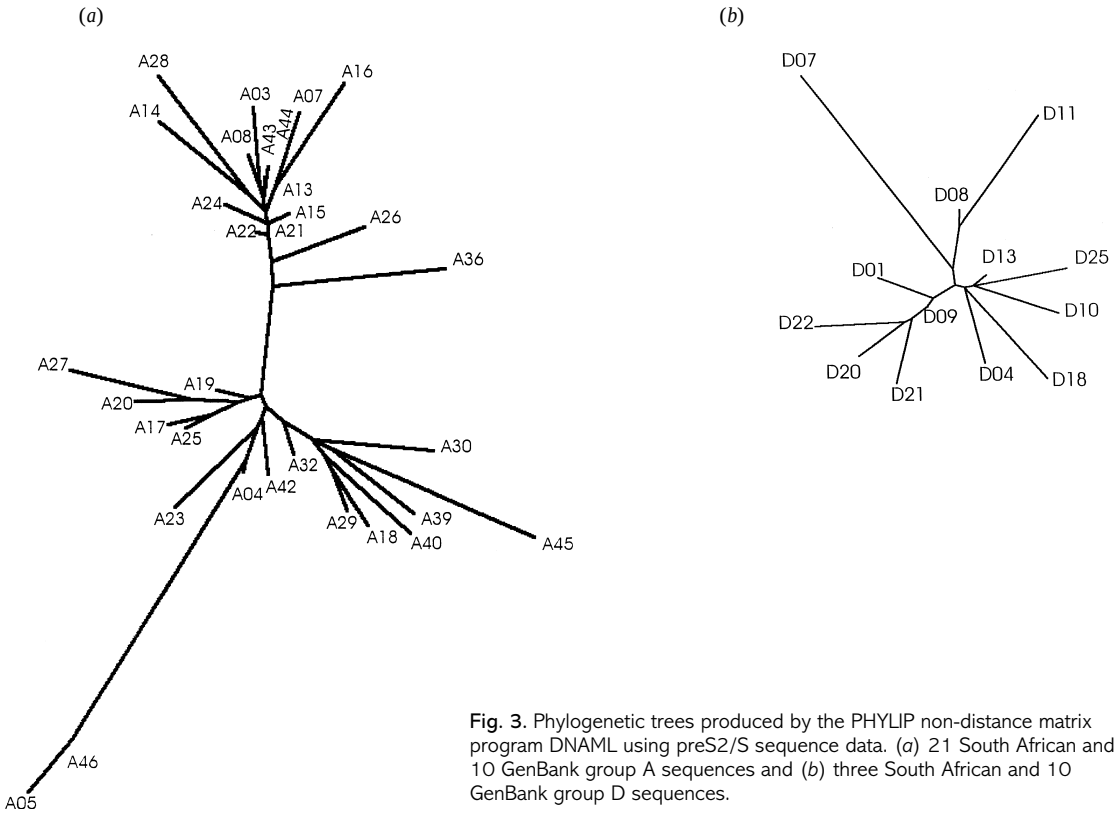

Fig. 3. Phylogenetic trees produced by the PHYLIP non-distance matrix program DNAML using preS2/S sequence data. (*a*) 21 South African and 10 GenBank group A sequences and (*b*) three South African and 10 GenBank group D sequences.

(*a*)

	A	B	C	D	E	F
A	**4.70** / **3.11**	13.09	12.32	11.55	10.70	20.23
B	7.38	**2.64** / **1.78**	15.22	14.28	13.08	23.71
C	8.07	8.05	**3.82** / **2.04**	11.22	9.72	21.38
D	6.60	7.38	6.86	**4.17** / **2.57**	9.05	20.70
E	6.83	7.47	6.41	5.83	**3.71** / **2.27**	18.57
F	11.02	11.14	11.06	10.10	9.86	**1.39** / **1.41**

(*b*)

	A	A'
A	**2.91** / **1.73**	6.07
A'	3.81	**4.74** / **2.95**

Fig. 4. Mean intra- and intergroup percentage divergence in nucleotide and amino acid sequences of the preS2/S gene region of (*a*) the six genotypic groups A toF and (*b*) the two subgroups of A. The intragroup percentage difference is boxed. The nucleotide divergence is shown on the left, and the amino acid divergence is shown on the right of the diagonal line.

A Lys to Arg switch at amino acids 122 and 160 of the S gene defines the *d/y* and *w/r* serotype specificities, respectively (Okamoto *et al.*, 1987, 1989). Ile^{126} is thought to abolish the *w* determinant and thus define the *r* determinant (Norder *et al.*, 1992). No other sites previously regarded as *d/y* or *w/r* determinants have consistently maintained their specificity as global sequence data accumulate. Residues 127 and 134, however, did consistently discriminate between the *w* determinants in the specimens examined. As previously described, Pro^{127} indicates *w1* or *w2*, Thr^{127} signifies *w3* and, since both groups E and F have Leu^{127}, without ambiguity, this amino acid most probably determines *w4*. *w2* specificity can be discriminated from *w1* by examining amino acid 134 where Phe^{134} is indicative of the former and Tyr^{134} of the latter. Based on these sites, all our group A *adw* specimens were of specificity *w2*.

The *ay* specificity was detected in six of the 30 southern African specimens in this study. These included the three South African group D specimens (namely, D20, D21 and D22) as well as the two group A specimens, A39 (*ayw*) and A40 (*adyw*), and the Zimbabwean specimen, A45. This 20% occurrence is 8·6 times higher than the previously reported figure of 2·3% from serotype analysis from southern Africa (Couroucé-Pauty & Soulier, 1983).

Variations specific to a particular genotypic group tend to cluster together in the same region of the genome. From Fig. 5, it can be seen that group B-specific changes occur at amino

	aa	A44	A43	A03	A07	A13	A08	A02	A06	A11	A12	A14	A28	A15	A16	A21	A36	A22	A24	A09	A01	A26	A10	A23	A20	A19	A17	A25	A27	A39	A29	A18	A32	A40	A30	A31	A34	A38	A04	A42	A45	A05	A46	A41	B01	B02	B03	B04	B05	B06	BO7	E01	E02	E03	E04	E06	E05
Pre-S2 Region	7	A	-	-	-	-	-	-		-	-	R	R	-	-	-	-	-	-			-	-	T	-	T	-	-	-	-	-	-	-	-	X	-		-	-	-	-	-	-		T	T	T	T	T	T		T	T		T	T	T
	10	Q	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	K	-	R	K	R	K	-	-	-	-	X	-		-	-	-	-	-	-		-	-	-	-	-	-		-	-		-	-	-
	11	T	A	A	A	A	A	A		A	A	A	A	A	A	A	A	A	A			A	A	A	V	A	A	A	A	A	A	A	A	A	X	A		A	A	A	A	A	A		-	-	-	-	-	-		A	A		A	A	A
	13	Q	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	K	X	-		-	-	-	-	-	-		-	-	-	-	-	-		-	-		-	-	-
	19	G	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	X	-		-	-	-	-	-	-		A	A	A	A	A	A		-	-		-	-	-
	22	L	-	-	F	F	-	F		F	-	F	-	F	-	F	F	F	F			F	F	F	-	F	F	-	F	F	F	-	F	F	X	F		F	F	F	F	F	F		F	F	F	F	F	F		F	F		F	F	F
	31	T	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		R	-	-	-	I	I		-	-	-	-	-	-		-	-		-	-	-
	32	V	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	L	L	L	L	L	L	L	L	H	H	L	L	L	L		L	L	L	-	-	-		-	-	-	-	-	-		-	-		-	-	L
	35	A	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	V	V	V	V	V	V	V	V	V	V	V	V	V	V		V	V	V	V	V	V		V	-	V	-	-	V		V	V		V	V	V
	36	P	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-		Q	Q	Q	Q	Q	Q		-	-		-	-	-
	37	N	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	T	-	T	T		-	-	-	-	-	-		T	T		T	T	-
	38	I	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	T	-	T		T	-	-	-	-	-		T	T	T	T	T	T		T	T		T	T	T
	39	A	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-		-	V	V	V	-	-		-	-		-	-	-
	41	H	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-		S	A	S	A	A	S		L	L		L	L	-
	42	I	-	-	-	-	-	-		-	-	-	-	-	M	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-		-	-	-	-	-	-		-	-		-	-	-
	46	S	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	F	F		L	L	L	L	L	L		F	F		F	F	F
	47	A	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	S	S	S	S	S	S	S	S	S	S	S	S	S	S		S	S	S	S	S	S		S	S	S	S	S	S		S	S		S	S	S
	48	R	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	T	T	T	T	T	-	T	K	-	-	-	-		-	-	-	-	-	-		T	K	K	K	K	T		-	-		-	-	T
	49	T	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	I	-	I	I		-	-	-	-	-	-		I	I		I	I	-
	51	D	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-		-	-		-	-	-
	53	V	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	A	A	-	-	-	-	-	-	A	A	A	A	A	A	-	-	A	A	A	A	A		-	-	-	-	-	-		A	A		A	A	A
	54	T	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-			-	P	P	P	P	P	P	P	S	L	L	S	L	A	L	P	P	-	P	S	P	P		P	P	P	P	P	P		P	P		P	P	P
Surface Region	2	E	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	3	N	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			S	S	S	-
	4	I	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	5	T	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	A	A	A	A	A		A			-	-	-	-
	8	F	-	-	-	-	-		-			-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	S	-	-	-	L	L	L	L	L		L			-	-	-	-
	14	V	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	A	L	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	18	G	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	19	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	24	R	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	K	K	K	K	K		K			K	K	K	K
	45	S	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	A	-	V	V	-	T	T	T	T	T		T			A	A	A	A
	46	P	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	47	V	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	49	L	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	P	P	-	-	-	-	-	-		-			-	-	-	-
	56	P	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	Q	Q	Q	Q	Q		Q			-	-	-	-
	57	T	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	I	I	I	I	I		I			-	-	-	-
	59	N	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	S	S	S	S	S		S			-	-	S	-
	61	S	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	64	S	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	C	C	C	C	C		C			-	-	-	-
	68	I	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	T	T	-	-	-	-	-	-		-			-	-	-	-
	85	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	C	C	C	C	S		C			-	-	-	-
	110	I	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	113	S	-	-	-	-	-		-			-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	T	-	-	-	-	-		-			-	-	-	-
	114	T	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	S	S	-	S	S	S	S	L		S			S	S	S	S
	122	K	-	-	-	-	-		-			-	F	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	R	-	-	-	R	-	-	-		-	-	R	-	-	-	-	-	-	-	-		R			R	R	R	-
	126	T	-	-	-	-	-		-			I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	A		-			-	-	-	-
	127	P	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			L	L	L	-
	131	N	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	T	T	T	T	T		T			T	T	T	T
	134	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	Y	Y	-	-	-	-	-	-		-			-	-	-	I
	140	T	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			S	S	S	-
	143	T	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	M	S	S	S	-	M	-	-	-		-			S	S	S	S
	158	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	K	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	159	A	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			G	G	G	-
	160	K	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	161	Y	-	-	-	-	-		-			-	-	-	-	F	-	-	-	-	-	-		F	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			F	F	F	F
	168	V	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-		-			A	A	A	-
	178	P	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	183	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	184	V	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			A	A	A	A
	193	S	-	L	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	L
	194	A	-	-	-	-	-		-			-	-	-	-	-	-	-	-	V	V	-		-	-	-	-	-	-	V	-	V	V	-	-		-		-	-	V	-	-	-	V	V	V	V	-		V			V	V	V	-
	198	M	-	-	-	-	-		-			-	I	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	I	-		I			-	-	-	-
	204	S	-	-	-	-	R		-			-	-	-	-	-	-	-	-	N	-	-		R	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	N
	206	Y	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	207	S	-	-	-	-	-		-			-	N	-	-	-	-	-	-	N	N	N		N	N	N	N	N	N	N	N	N	-	N	N		N		N	N	N	N	N	-	N	N	N	N	N		N			N	N	N	N
	209	V	-	-	-	-	-		-			-	-	-	-	-	-	-	-	L	L	L		L	L	L	L	L	L	L	L	L	-	L	L		-		L	L	L	L	L	-	L	L	L	L	L		L			L	L	L	L
	210	S	-	R	-	-	-		-			R	-	-	-	-	R	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	R		-			-	-	-	-
	213	I	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	T	-	-	-	M	M	M	M	-		M			-	-	-	-
	220	F	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-
	225	Y	-	-	-	-	-		-			-	-	-	-	-	-	-	-	-	-	-		I	-	-	-	-	-	-	-	-	-	-	-		-		-	-	-	-	-	-	-	-	-	-	-		-			-	-	-	-

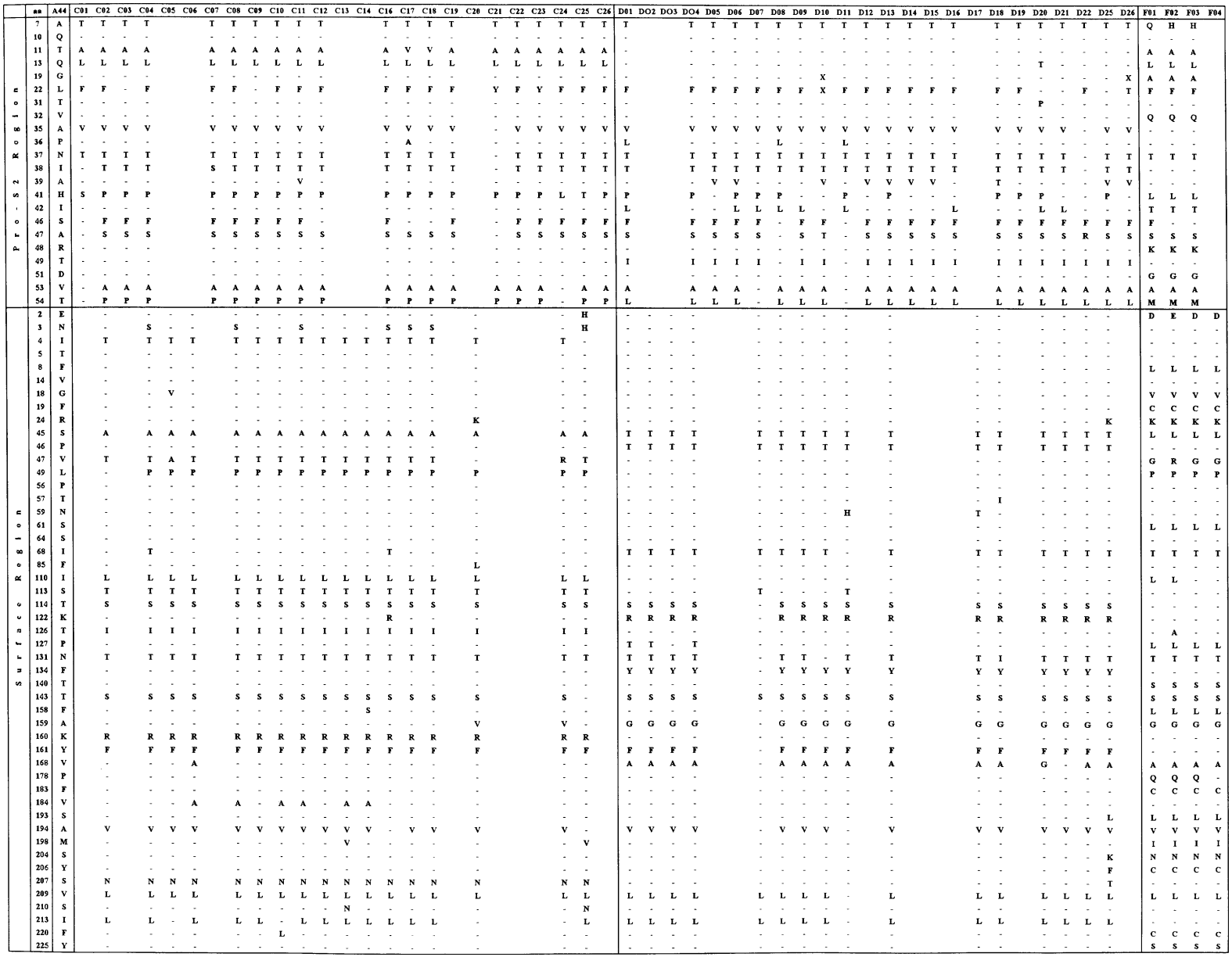

Fig. 5. PreS2/S amino acid present at HBV variant sites in representative specimens from groups A to F. The amino acids are numbered in ascending order from amino acid 1 to 55 for the preS2 region and from amino acid 1 to 226 for the S region.

acids 36 and 46 in the preS2 region and at amino acids 5, 56, 57, 64, 85 and 213 in the S region. Group F-specific sites occur at amino acids 32, 42, 51 and 54 in the preS2 region and at 18, 19, 45, 61, 158, 178, 183, 206, 220 and 225 in the S region, which are all in the fourth helix or amino-terminal region of the major surface antigen. Asn^{204}, previously defined as a group F determinant (Norder *et al*., 1994), was present in two non-F isolates, A09 and E05 (Fig. 5). Also, group C determinants occur at surface amino acid 4 and at two sites, amino acids 126 and 160, which define the *r* determinant in the extracellular loop. In this same region, group A is identified by residue 45. Group D was defined in the S region by Thr^{45}, Thr^{46} and Tyr^{134}. Group E did not have any definitive sites but shared three sites with group F, preS2 site 41 and surface antigen sites 127 and 140.

Leu^{22} shows considerable variation and forms part of the immunodominant region extending from preS2 amino acids 12 to 24, which are reported to be involved in the primary antibody response (Meisel *et al*., 1994). As shown in Fig. 5, Thr^{11} (in groups B and D) and Leu^{13} (in groups C and F) occur in a preS2 region, amino acids 1 to 15, which is both HLA class I and class II restricted (Chisari, 1995) and the variable residues 7 and 22 are at T cell receptor contact sites.

A series of variations, predominantly in the South African group A sequences, which tended to cluster in the preS2 region, set A′ apart from the rest of group A, which is otherwise well conserved (Fig. 3*a*). Also grouping with these unique South African specimens was the Zimbabwean isolate (A45), the French isolate (A04) and clone pFDW294 (A42) from the Philippines, previously observed to be most similar to the Zimbabwean isolate. When the preS2 sequences from Kidd-Ljünggren *et al*. (1994) and the S only sequences from Norder *et al*. (1992) are included in the analyses the new subgroup expands to 23 (including 66·7% of the southern African specimens); 19 of these (82·6%) are African in origin since the Mozambican PLC/PRF/5 hepatoma cell line sequence as well as a sequence from Cameroon and the three incomplete South African sequences cluster with this group. Of interest, the only other addition to the group is Kidd-Ljünggren's specimen LS (A10) from the Philippines. Additions to the other subgroup include three European specimens (A06, A11 and A12), a Taiwanese specimen (A41) and Lo's specimen (A02) which is of Far Eastern origin.

In the phylogenetic trees (Figs 2 and 3) some specimens such as the mutant A05 and recombinant A46, referred to by Tran *et al*. (1991), as well as the Zimbabwean variant A45 show considerable variation from the rest of the specimens. A46 and A05 show a series of common mutations setting them apart from the rest of subgroup A′ at preS2 sites 31, 37 and 49 and surface antigen sites 45, 49, 68, 114, 134 and 143. D07, the GenBank HPBAYWCI mutant (Lai *et al*., 1991), which was isolated from a patient with HBsAg negative chronic liver disease, behaves like a putative recombinant having group A preS2 sequences and group D surface antigen sequences. This was revealed by the fact that it falls into 'significantly discordant positions in different trees' (Hahn *et al*., 1995), grouping with group D when the S region alone was analysed but with group A when the preS2 region alone was used. When the full preS2/S sequence was used in the analysis, D07 becomes an outlier separated from known groups (Fig. 2).

In the preS2 region, five of the South African specimens show a change at the T cell receptor contact site at amino acid 7. In two of the specimens Ala^{7} is changed to Arg^{7} and in the three others it is changed to Thr^{7}. This is not, however, a subgroup-specific change as the former pair of specimens are not part of the new subgroup and Thr^{7} was also used previously as a determinant for group D. At amino acid 10, Gln is changed to Lys or Arg in five of our specimens and amino acids 7 to 22 are deleted in another. This site is subgroup-specific as the variants are all part of a subdivision of A′. This section (A19, A27, A20, A17 and A25, see Fig. 3) is also represented by changes at amino acids 48, 53 and 54. PreS2 amino acids 32, 35, 47 and 54 and surface antigen residues 207 and 209 split segment A′ from the rest of group A. A01 (Cameroon) and A09 (Mozambique) have these latter two A′ determinant sites. Amino acids 47 (Ala to Ser) and 48 (Arg to Thr/Lys), which define group A′, were part of the genotype-dependent recognition site at the carboxy terminus of the preS2 sequence, amino acids 42 to 49 (Meisel *et al*., 1994).

There are indications (e.g. group B preS2 amino acids 33, 35, 39, 41 and 48) that other groups might show subdivisions when further group members are sequenced. Certainly, an understanding of highly conserved sites (e.g. amino acids 86 to 98, a region containing two HLA-A2 epitopes), as well as highly variant regions may be helpful in understanding the virus–host interactions and the form of disease manifestation in different individuals following HBV infection. The distribution of mutational hot spots in the alternate reading frame of the polymerase gene could also help to determine those portions of the P gene with essential functions (Fig. 1.)

The first two authors contributed equally to this project and should be regarded as equal first authors. The work at the NIV was supported by the Poliomyelitis Research Foundation.

References

Antoni, B. A., Rodriguez-Crespo, I., Gomez-Gutierrez, J., Peterson, D. & Gavilanes, F. (1994). Site-directed mutagenesis of cysteine residues of hepatitis B surface antigen: analysis of two single mutants and the double mutant. *European Journal of Biochemistry* **222**, 121–127.

Ashton-Rickardt, P. G. & Murray, K. (1989). Mutants of the hepatitis B virus surface antigen that define some antigenically essential residues in the immunodominant 'a' region. *Journal of Medical Virology* **29**, 196–203.

Chirara, M. M. & Chetsanga, C. J. (1994). Variant of hepatitis B virus isolated in Zimbabwe. *Journal of Medical Virology* **42**, 73–78.

Chisari, F. V. (1995). Hepatitis B virus immunopathogenesis. *Annual Review of Immunology* **13**, 29–60.

Couroucé, A.-M., Holland, P. V., Muller, J. Y. & Soulier, J. P. (1976). HBs antigen subtypes. *Bibliotheca Haematologica* **2**, 1.

Couroucé-Pauty, A.-M., Lemaire, J. M. & Roux, J. F. (1978). New hepatitis B surface antigen subtypes inside the ad category. *Vox Sanguinis* **35**, 304–308.

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

Estacio, F. 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 three reported genomes of the same subtype. *Journal of Gastroenterology and Hepatology* **3**, 215–222.

Felsenstein, J. (1993). PHYLIP: phylogeny inference package (Version 3.5c). Distributed by the author. Department of Genetics, University of Washington, Seattle, Washington, USA.

Fujii, H., Moriyama, K., Sakamoto, N., Kondo, T., Yasuda, K., Hiraizumi, Y., Yamazaki, M., Sakaki, Y., Okochi, K. & Nakjima, E. (1992). Gly145 to Arg substitution in HBsAg of immune escape mutant of hepatitis B virus. *Biochemical and Biophysical Research Communications* **184**, 1152–1157.

Hahn, B. H., Robertson, D. L. & Sharp, P. M. (1995). Intersubtype recombination in HIV-1 and HIV-2. In *Human Retroviruses and Aids*, pp. III-22–III-29. Edited by G. Myers, B. H. Hahn, J. W. Mellors, L. E. Henderson, B. Korber, K.-T. Jeang, F. E. McCutchan & G. N. Pavlakis. New Mexico: Los Alamos National Laboratory.

Harrison, T. J., Oon, C.-J. & Zuckerman, A. R. (1991). Independent emergence of a vaccine-induced escape mutant of hepatitis B virus. *Journal of Hepatology* **13**, S105–107.

Kidd-Ljünggren, K., Couroucé, A.-M., Oberg, M. & Kidd, A. H. (1994). Genetic conservation within subtypes in the hepatitis B virus pre-S2 region. *Journal of General Virology* **75**, 1485–1490.

Lai, M. E., Melis, A., Mazzoleni, A. P., Farci, P. & Balestrieri, A. (1991). Sequence analysis of hepatitis B virus genome of a new mutant of ayw subtype isolated in Sardinia. *Nucleic Acids Research* **19**, 5078.

Lauder, I. J., Lin, H.-J., Lau, J. Y. N., Siu, T.-S. & Lai, C.-L. (1993). The variability of the hepatitis B virus genome: statistical analysis and biological implications. *Molecular Biology of Evolution* **10**, 457–470.

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

Meisel, H., Sominskaya, I., Pumpen, P., Pushko, P., Borisove, G., Deepen, R., Lu, P., Spiller, G. H., Kruger, D. H., Grens, E. & Gerlich, W. H. (1994). Fine mapping and functional characterization of two immunodominant regions from the preS2 sequence of hepatitis B virus. *Intervirology* **37**, 330–339.

Norder, H., Hammas, B., Lofdahl, S., Couroucé, 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., Hammas, B., Lee, S.-D., Bile, K., Couroucé, A.-M., Mushawar, 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, structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* **198**, 489–503.

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

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R., Imai, M., Miyakawa, Y. & Mayumi, M. (1989). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *Journal of General Virology* **69**, 2575–2583.

Oon, C.-J., Lim, G.-K., Ye, Z., Goh, K.-T., Tan, K.-L., Yo, S.-L., Hopes, E., Harrison, T. J. & Zuckerman, H. J. (1995). Molecular epidemiology of hepatitis B virus vaccine variants in Singapore. *Vaccine* **13**, 699–702.

Rivkina, M. B., Lunin, V. G., Mahov, A. M., Tikchonenko, T. I. & Kukain, R. A. (1988). Nucleotide sequence of integrated hepatitis B virus DNA and human flanking regions in the genome of the PLC/PRF/5 cell line. *Gene* **64**, 285–296.

Tran, A., Kremsdorf, D., Capel, F., Housset, C., Dauguet, C., Petit, M.-A. & Brechot, C. (1991). Emergence of and takeover by hepatitis B virus (HBV) rearrangements in the pre-S/S and pre-C/C genes during chronic HBV infection. *Journal of Virology* **65**, 3566–3574.

Waters, J., Kennedy, M., Voet, P., Hauser, P., Petre, J., Carman, W. & Thomas, H. C. (1992). Loss of a common 'a' determinant of hepatitis B surface antigen by a vaccine escape mutant. *Journal of Clinical Investigation* **90**, 2543–2547.

Whittle, H. C., Inskip, H., Hall, A. J., Mendy, M., Downes, R. & Hoares, S. (1991). Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia. *Lancet* **337**, 747–750.

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 31 October 1996; Accepted 5 March 1997