View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
ScienceDirect

Virology 365 (2007) 113 – 124

VIROLOGY

www.elsevier.com/locate/yviro

High level of genetic heterogeneity in S and P genes of genotype D hepatitis B virus

Chiara De Maddalena [a], Camilla Giambelli [a], Elisabetta Tanzi [b], Daniela Colzani [b], Monica Schiavini [c], Laura Milazzo [a], Flavia Bernini [a], Erika Ebranati [a], Antonietta Cargnel [c], Raffaele Bruno [d], Massimo Galli [a], Gianguglielmo Zehender [a,*]

[a] *Department of Clinical Sciences Luigi Sacco, Section of Infectious Diseases, University of Milan, Milan, Italy*
[b] *Department of Public Health–Microbiology–Virology, University of Milan, Milan, Italy*
[c] *II Department of Infectious Diseases, Luigi Sacco Hospital, Milan, Italy*
[d] *Division of Infectious and Tropical Diseases, IRCCS S. Matteo Hospital, University of Pavia, Pavia, Italy*

Received 27 December 2006; returned to author for revision 30 January 2007; accepted 9 March 2007
Available online 23 April 2007

Abstract

The genetic heterogeneity of hepatitis B virus (HBV) genotypes and subgenotypes was investigated by directly sequencing amplified PreS, S and P genes of HBV isolates obtained from the plasma of 99 subjects with chronic HBV infection. Genotype D showed the greatest intragenotypic and intrasubgenotypic divergence: in particular, the *a determinant* was mutated in 58.2% of the genotype D patients, two of whom showed prototypic vaccine-induced escape mutants at codon 145. Moreover, five sites under significant positive selection were found in the S protein of the D isolates: one in the *a determinant* and four in the highly hydrophobic C terminal.

Our results suggest that careful surveillance of vaccine-induced escape mutants should be considered in populations with highly frequent genotype D infections, and raise questions concerning the possible relationship between the genetic heterogeneity, host immunity and pathogenicity of this HBV genotype.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Hepatitis B virus; Genetic variability; Molecular evolution; Genotypes; Natural selection; Genetic heterogeneity; Pathogenesis

Introduction

Hepatitis B virus (HBV) is a partially double-stranded DNA virus of the Hepadnaviridae family that infects an estimated 300 million people worldwide. What makes it a unique DNA virus is its use of an intermediate RNA and a reverse transcriptase (rt) step during replication, which is why its evolutionary rate of between 1.4 and 4.6×10^{-5} substitutions per site per year is higher than that of other DNA viruses (Okamoto et al., 1987; Orito et al., 1989).

Significant variability in HBV was first recognised in the early 1970s, when the existence of mutually exclusive determinants (d/y and w/r) coupled with a more conserved domain (*a determinant*) were serologically defined at surface antigen level, and a number of viral serotypes were identified (Kidd-Ljunggren et al., 2002). The availability of a number of complete sequences of HBV isolates has led to the identification of eight genotypes (A–H) largely corresponding to the known serotypes which, more recently, have been further classified into subgenotypes with different geographic and ethnic distributions (Norder et al., 2004).

HBV genotypes have also been associated with different clinical outcomes (Schaefer, 2005), with genotypes D and F being associated with more severe disease progression and a worse prognosis than genotype A (Sanchez-Tapias et al., 2002; Thakur et al., 2002), although partially contrasting results have been reported by other authors (Chattopadhyay et al., 2006; Kumar et al., 2005).

A further correlate of the high degree of variability of HBV is the selection of point mutations in different viral

* Corresponding author. Dipartimento di Scienze Cliniche "L. Sacco", Sezione Malattie Infettive, Università di Milano, c/o Ospedale L. Sacco, Via G.B. Grassi 74, 20157 Milano, Italy. Fax: +39 0250319768.
E-mail address: gianguglielmo.zehender@unimi.it (G. Zehender).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.03.015

genes under host immunity or drug treatment. In particular, the immune response solicited by natural infection or vaccination can induce the emergence of mutants in the S gene which, particularly those in the most immunogenic *a determinant*, are positively selected by their ability to evade immune defences (Zuckerman and Zuckerman, 2003).

The aim of this study was to investigate the genetic heterogeneity of S and P genes in the different HBV genotypes and subgenotypes, and the prevalence of mutations at nucleotide and amino acid level, particularly those in the common neutralising epitope (the *a determinant*) of the HBV surface antigen and viral reverse transcriptase. We also investigated the role of selection in driving the heterogeneity of the different HBV genotypes.

Results

Population characteristics

Genotype D was the most represented in our case file (55 cases, 55.5%), followed by genotype A (30 cases, 30.3%). Genotypes F and G accounted for respectively seven (7.1%) and four (4.1%) cases, and genotypes B, C and E for one case each (1.0%). Thirty-nine patients (39.4%) were co-infected with HIV-1. Among the HIV-1-positive individuals, 15 (38.5%) were infected by genotype A, 18 (46.1%) by genotype D, four (10.2%) by genotype G, and one each (2.6%) by genotypes B and E.

A complete phylogenetic analysis was made on the whole S region of the 99 isolates, amplified as reported in Table 1. The S sequences were 678 nucleotides long (226 aa) in all but two of the isolates, which were 651 and 654 nct because of deletions in positions 69–95 and 318–341 (aa 23–32 and aa 106–114). The length of the PreS regions varied from 342 (in 93/99, 94%) to 285 nct, because of deletions affecting PreS1 (nct: 181–230 and 192–224) in two genotype D isolates, and PreS2 (the deletion of 1–20 residues between nct 1 and nct 72) in one genotype A and five genotype D isolates.

The phylogenetic analysis of the 99 isolates (Fig. 1) showed seven well-supported clades corresponding to the major HBV genotypes (A–G); their topology was confirmed by both NJ and ML. Two recombinant isolates were identified: one showing a shift between genotypes A (in Pre/S) and G (in gene S), and the other between genotypes D and A. Different clades supported by significant bootstrap values were identified within genotypes A, F and D. The genotype A isolates segregated into two significant clades (Fig. 1): the first included three isolates corresponding to A1, and the second 27 isolates corresponding to A2 subgenotypes. Likewise, the genotype F isolates formed two significant groups matching the F1 (3 sequences) and F2 subgenotypes (4 sequences). The genotype D isolates showed the most complex topology, with only partially significant clades corresponding to known subgenotypes D1 (5 isolates), D2 (8 isolates), D3 (40 isolates) and D4 (2 isolates) (Fig. 2).

Distance analysis of S and P genes

Comparison of the mean intergenotypic distances showed that strain F was the most divergent (the mean distances

Table 1
Primers for HBV–DNA env/pol amplification and sequencing

HBV–PCR	Primer name	NCT [a] position	Sequence (5′–3′)	Cycling conditions
Outer primers for 1st fragment: set number 1	HBV-3	2474–2497 (S [b])	GGGRAAYTTTACKGGGCTBTAYTC	45 cycles:
	HBV-15	663–646 (AS [c])	AAACGGRCTGAGGCCCMC	95 °C, 50 s
				57 °C, 50 s
				72 °C, 100 s
Inner primers: set number 3	HBV-7	2965–2984 (S)	ARTCCMGATTGGGACYWCAA	40 cycles:
	HBV-4	280–257 (AS)	TAGAAAATTGAGAGAAGTCCACCA	95 °C, 40 s
				54 °C, 40 s
				72 °C, 70 s
Inner primers: set number 4	HBV-5	180–203 (S)	TAGGACCCCTKCTCGTGTTACAGG	40 cycles:
	ERS-IN3	655–637 (AS)	TGAGGCCCACTCCCATAGG	95 °C, 40 s
				58 °C, 40 s
				72 °C, 70 s
Outer primers for 2nd fragment: set number 2	HBV-9	202–220 (S)	GGCGGKGTKTTTCTTGTTG	45 cycles:
	HBV-6	1108–1085 (AS)	AAGTTGGCGARAARRYRAAAGCCT	95 °C, 45 s
				55 °C, 45 s
				72 °C, 100 s
Inner primers: set number 5	ERS-IN2	409–428 (S)	CATCCTGCTGCTATGCCTCA	40 cycles:
	HBV-6	1108–1085 (AS)	AAGTTGGCGARAARRYRAAAGCCT	95 °C, 40 s
				58 °C, 40 s
				72 °C, 70 s

R: A/G; Y: C/T; K: G/T; B: G/T/C; M: A/C; W: A/T.
[a] NCT: nucleotide.
[b] S: sense primer.
[c] AS: antisense primer.

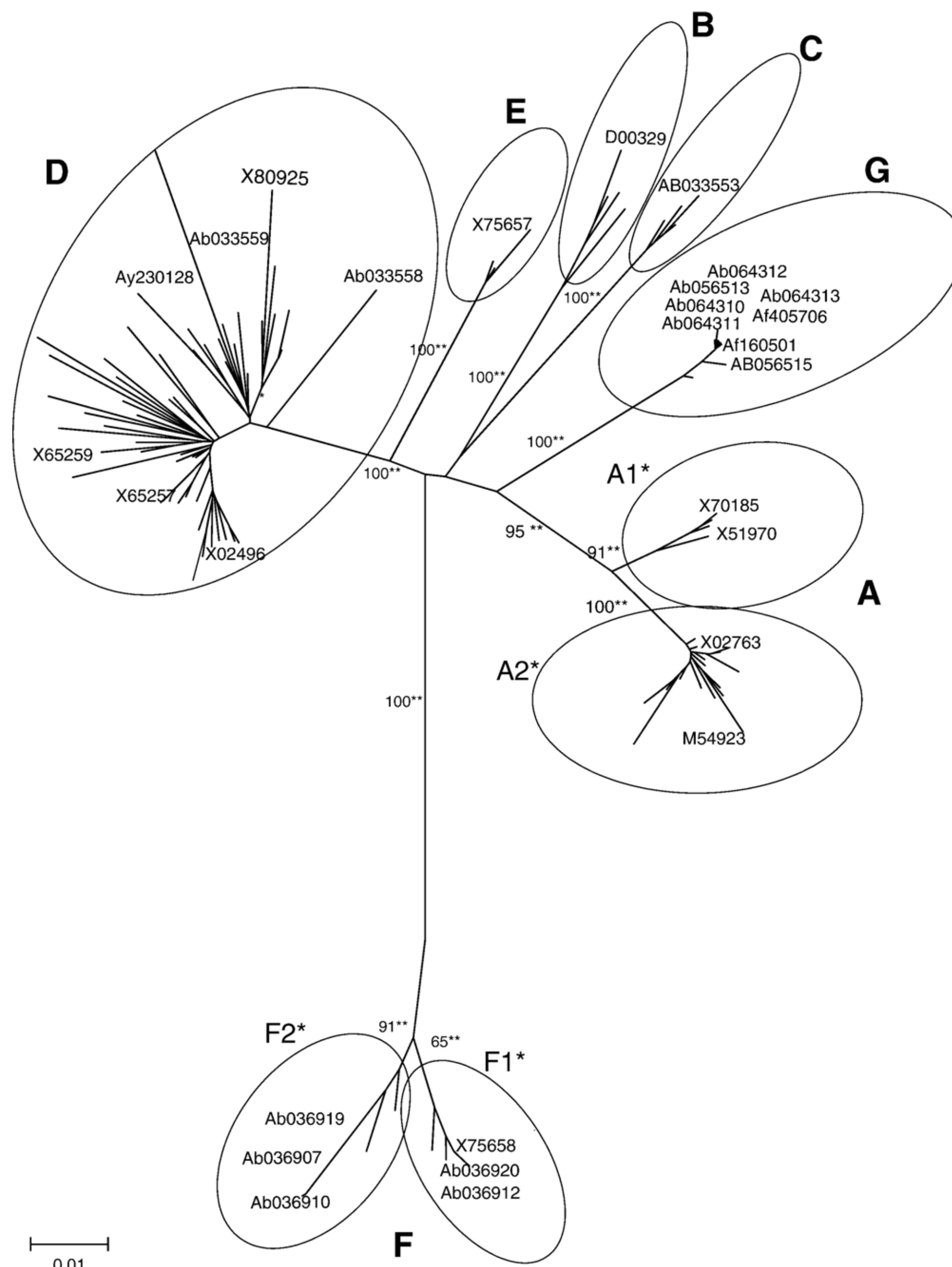

Fig. 1. Unrooted NJ tree of 99 patient isolates with 28 reference sequences (indicated by accession numbers) using all sequences 1182 nucleotides long. The numbers on the branches denote bootstrap values (percentages) at NJ analysis, and the asterisks the level of statistical significance of the branch length at ML analysis (*$0.05 \leq p > 0.001$ and **$p \leq 0.001$; see Materials and methods for details). The A and F subgenotypes are surrounded by ellipses.

between F and the other genotypes varied from 0.084 [0.012] to 0.097 [0.013] in S sequences versus the mean distances from 0.047 [0.009] to 0.061 [0.008] between the other genotypes).

Table 2 shows the mean distances between isolates of the same genotype (intragenotypic distance – IG) or subgenotype (intrasubgenotypic distance – ISG). The mean intragenotypic divergence of the genotype D isolates was significantly greater than that of genotypes A and G in all of the studied portions of the viral genome. In particular, the *a determinant* was highly conserved in genotypes A and G (mean distances 0.9 [0.4] and 0 substitution/10^2 sites), but showed a divergence of 4.1 [0.9] in the D strains (Table 2, $p < 0.001$ by the Mann–Whitney test). The distance between the strains of genotype F was always less than that observed in genotype D, but the difference was statistically significant only in the case of the *a determinant* ($p = 0.006$).

The mean intrasubgenotypic distances of the most represented D subgenotypes (D1, D2 and D3) were always greater than those of the A and F subgenotypes (Table 2). In particular, comparison of the intrasubgenotype divergences of the two most represented groups showed that D3 (40 isolates) had higher values than A2 (27 isolates) in all of the analysed regions ($p < 0.001$ by Mann–Whitney's test).

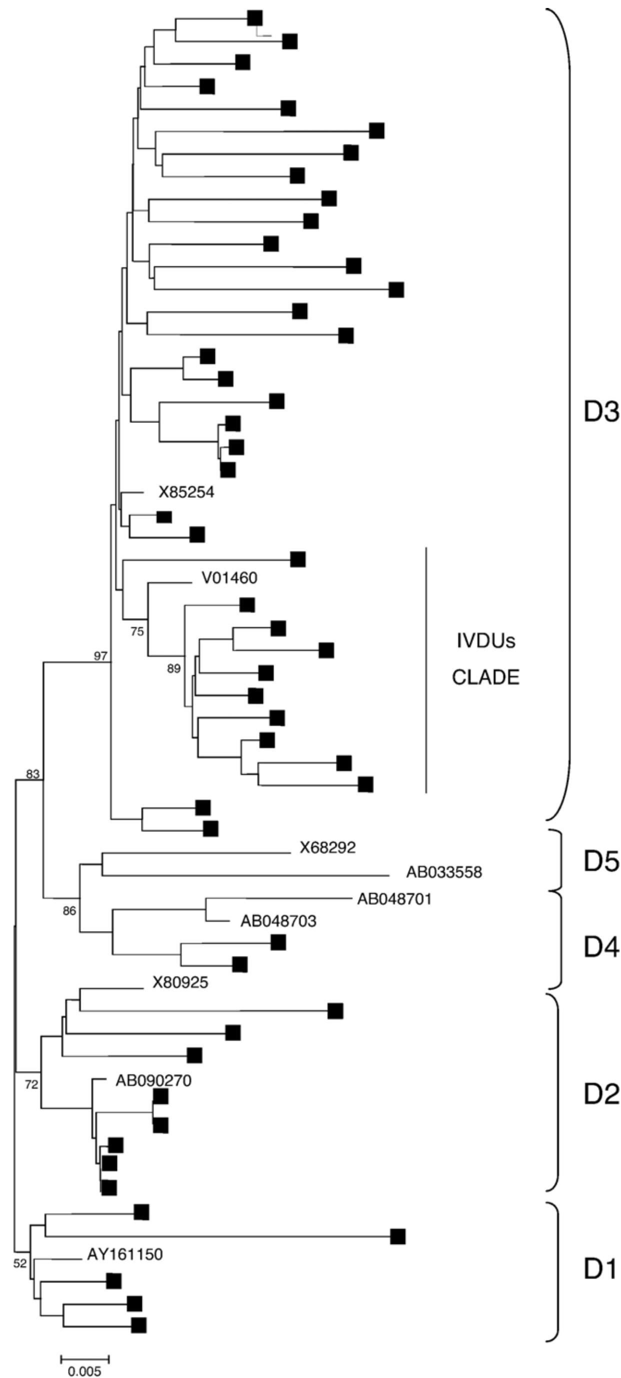

Fig. 2. Unrooted NJ tree of 55 genotype D isolates using the entire 1182-nucleotide fragment. The numbers on the branches denote bootstrap values (percentages) at NJ analysis. The reference sequences can be recognised based on their accession numbers. Significant clades have been highlighted.

There was no association between the greater heterogeneity of genotype D and HIV-1 infection, although the coinfected subjects had the highest intragenotypic divergence values in the S (2.7 sub/10^2 sites vs. 2.2 sub/10^2 sites) and *a determinant* regions (5.1 sub/10^2 sites vs. 3.2 sub/10^2 sites).

Table 3 shows the mean synonymous (d_S) and non-synonymous substitution rates (d_N) for the S, *a determinant*, PreS and P regions, estimated using the Nei–Gojobori model: d_S was always higher than d_N regardless of the region (mean d_N/d_S ratio 0.4 in the S and P genes, and 0.2 in the *a determinant* and PreS2). Genotype D always had the highest d_N, which was about three times higher than that of genotype A strains in the whole S coding region and 29 times higher in the *a determinant*. Similar results were obtained when intrasubgenotype d_S and d_N were compared (data not shown).

PreS/S and P proteins

At protein level, the prevalence of isolates carrying at least one amino acid mutation was 81.8% (81/99 isolates) in the small surface antigen, 39.4% (39/99) in the *a determinant*, 56.7% (55/97) in the PreS proteins, and 92.8% (90/97) in the polymerase.

The frequency of mutated strains in the S protein was always more than 70% for genotypes D, A and F, and similar mutant frequency was observed in all of the most represented subgenotypes (Table 4); only genotype G did not show any variation in this region. Limiting the analysis to the *a determinant* region among the subjects infected with genotype D, the mean prevalence of those carrying mutated strains was 58.2%, as against 6.7% in those carrying genotype A ($p<0.001$ by the χ^2 test) and 0% in those carrying genotype G ($p<0.02$ by the χ^2 test). Likewise, comparing the two most numerous subgenotypes, the frequency of mutations in the *a determinant* was 55% in the D3 strains and 7.4% in the A2 isolates ($p<0.001$ by the χ^2 test).

Figs. 3A/B shows the amino acid mutations affecting more than 10% of the isolates, as deduced from the genotype A and D PreS/S sequences. All of the substitutions and affected amino acids are described in detail in the Supplementary Tables S1 and S2. The most frequent surface antigen mutation in the D strains was the substitution of S at site 207, which affected a mean 34.5% of D isolates and 32.5% of the D3 strains. Four mutations were highly prevalent among strains infecting HIV1-positive drug addicts (all D3): an S to T substitution in position 90 of PreS1 (prevalence among drug addicts 7/9, 77.8%); T31A in PreS2 (6/9, 66.7%); T125M in S protein (7/9, 77.8%); and P127T/L/A (8/9, 88.9%), which was also but less frequently present among the non-IVDUs (8/31, 25.8%). These mutations were frequently associated with each other: nine isolates (seven from IVDUs, one from a subject at risk for sexual transmission, and one without known risk) carried a variant characterised by three or four of these substitutions. All of these subjects were HIV-1-positive.

Known escape mutants were found in two patients, both carrying genotype D (2/55, 3.6%) and subgenotype D3 (2/40, 5%): one isolate was characterised by the G to A substitution at amino acid residue 145 (G145A), and the other showed the prototypic G to R substitution at the same position (G145R).

Of the 32 genotype D isolates mutated in the *a determinant*, 16 (50%) showed mutations only in the first loop, 1 (3%) only in the second loop, and 10 (31.2%) in both.

Figs. 3C/D shows the mutations in the polymerase protein affecting more than 10% of the genotype A and D isolates (see

Supplementary Tables S3 and S4 for detailed information concerning the individual substitutions).

The most frequent mutations affecting the viral polymerase included two associated with resistance to lamivudine: at site rt204 (the typical M to V/I substitution in the YMDD motif) and rt180 (shift from L to M). At the time of enrolment, 27 patients had been treated with lamivudine for HIV-1 infection for more than six months. Sixteen of them (59.2%) had the rtM204V/I mutation; the proportion of resistant strains was 81.8% (9/11) among the subjects with genotype A and 58.3% (7/12) in those with genotype D, whereas none of the four subjects carrying other genotypes (two with genotype G and two with genotype F) had mutant strains in the YMDD motif.

Site-specific selection pressure in S and P proteins

After excluding the isolates showing recombination, deletions or premature stop codons, the analysis was based on a total of 49 type D, 28 type A, three type G and seven type F isolates.

Model M2 (selection) was significantly better at the likelihood ratio test than M1 (neutral) only in the case of genotypes A and D, thus suggesting the presence of positive selection in these strains but not in the others (Table 5). No significant selection pressure was detected in the PreS coding region ($d_N/d_S = 0.32$ in genotype D, and 0.67 in genotype A).

In the S gene, the genotype D isolates showed five sites under positive selection at the 95% level of significance (5/226, 2.2% – Fig. 3A). In particular, one of the selected sites in the genotype D isolates was localised in the *a determinant*: the substitution of the Y in position 134 (present in all the D subgenotypes) by N or S or F. The other four were in the C-terminal of the S protein (S204N/R/T, Y206S/C/F/N, S207N/R/H/T and S210R/H/N). No sites under significant selective pressure were identified in the S coding region of A strains (Fig. 3B).

Analysis of the polymerase coding sequences (Figs. 3C/D) showed four sites under significant positive pressure in the genotype D isolates. There was one in the spacer (K106R/N) and three in the rt domain: rtY54D/H/N, rtQ215S/P/E/H (not included in the Fig. 3C because its frequency among the isolates was 7.7%, slightly less than 10%), and rtI266R/V (in the D1, D2 and D4 subgenotypes) or rtV266I/R (in the D3 subgenotype). All of the three sites under positive selection in the genotype A isolates were included in the RT protein, and two (rtM204: $\omega^2 = 5.5$, $p = 0.99$ and rtL180: $\omega^2 = 5.3$, $p = 0.97$) were associated with lamivudine resistance. Limiting the analysis to the patients on lamivudine treatment at enrolment, significant positive selection was also found in sites rt204 and 180 of the genotype D isolates ($\omega^2 = 2.9$, $p = 0.99$ and $\omega^2 = 2.8$, $p = 0.96$, respectively).

Because of the overlap of the S and P coding regions, six of the most frequent mutations in genotype D and three in genotype A affected both proteins (highlighted in the Fig. 3 by letters). We analysed codon rt215 of the P protein, and codon 207 of the S protein, with particular attention because both were under significant selection in genotype D. ML analysis showed $\omega = 5.96 \pm 1.27$ in codon 207, and $\omega = 3.69 \pm 0.49$ in codon rt215. Moreover, mutation rt215 was always accompanied by mutation 207, but only nine of the 19 strains with mutation 207 also had mutation rt215.

Interestingly, two contiguous codons of the RT region (I53V/S in genotype A and Y54D/H/N in genotype D isolates), both under selective pressure, overlapped on site 45 of the S region.

SLAC analysis of the PreS/S sequences showed that the greatest frequency of sites under purifying selection was in the PreS1 region (13 of the 59 studied sites, 22.0%) followed by PreS2 (7/55, 12.7%) and S protein (10/226, 4.4%). In strain A, only one amino acid was under significant negative pressure in the S protein (G44). All of the 30 negatively selected sites found in the P protein of genotype D (30/394, 7.6%), as well as the eight (8/394, 2%) found in genotype A, were included in the rt domain.

Discussion

Interest in studying HBV genotypes has increased since it has been shown that they can affect different infection outcomes (Kao et al., 2000; Sanchez-Tapias et al., 2002; Shiina et al., 1991a,1991b; Sumi et al., 2003; Thakur et al., 2002; Tsubota et al., 2001; Yuen et al., 2003) and treatment responses (Erhardt et al., 2005; Zollner et al., 2002), but the intrinsic or extrinsic factors explaining such influences are still unknown.

A number of observations suggest that the different genetic heterogeneity of viral genotypes could be one of these factors (Li et al., 1993; McMillan et al., 1996; Rodriguez-Frias et al., 1995).

Interestingly, we found that the mean intragenotypic distance of type D strains was significantly higher than that of types A and G in both PreS/S and P sequences.

The existence of subgenotypes with different ethno/geographic distributions has recently been demonstrated for genotypes A, B, C, D and F. In relation to the genotypes present in our population, three subgenotypes have so far been identified in genotype A (Kramvis et al., 2002; Makuwa et al., 2006), two in genotype F (Arauz-Ruiz et al., 1997), and five in genotype D (Banerjee et al., 2006; Norder et al., 2004). This last is the most widespread genotype, being highly prevalent in the Mediterranean area, Middle East, India, Indonesia, Australia and Papua. It has also been isolated in South Africa, Somalia and Alaska (Norder et al., 2004). It has been suggested that this geographic dispersion of genotype D is one of the reasons for its highly complex phylogenesis, and so we decided to extend the analysis of genetic variability to the subgenotypes. The mean distances of the D1, D2 and D3 subgenotypes (the most represented in our subjects infected with type D) were greater than those of all the A and F subgenotypes, particularly in the *a determinant* region, which suggests that the genetic heterogeneity of genotype D does not just depend on its phylogenetic complexity and is not characteristic of a specific geographic or ethnic subgroup.

As also suggested by the observation that mean d_N values were always higher in genotype and subgenotypes D, its high degree of polymorphism at nucleic acid level was accompanied

Table 2
Mean intragenotyipc (IG – bold) and intrasubgenotypic (ISG – italic) genetic distances (substitution/10^2 sites) and standard errors (SE) of the S, *a determinant*, PreS and P regions in the study isolates

Coding region	Mean intragenotype/intrasubgenotype distances (SE)											
	ISG				**IG**	*ISG*		**IG**	**IG**	*ISG*		**IG**
	D1	*D2*	*D3*	*D4*	**D**	*A1*	*A2*	**A** [a]	***G*** [a]	*F1*	*F2*	**F**
S	2.2 (0.4)	1.6 (0.3)	2.4 (0.3)	0.7 (0.3)	**2.5 (0.3)**	0.2 (0.1)	0.8 (0.2)	**1.0 (0.2)**	**0**	0.2 (0.2)	0.2 (0.1)	**1.3 (0.3)**
p [b]	–							*< 0.001*	*0*			*0.1*
a determinant	2.1 (0.8)	2.2 (0.8)	3.4 (0.7)	0.7 (0.7)	**4.1 (0.9)**	0.5 (0.5)	0.5 (0.5)	**0.9 (0.4)**	**0**	0	0	**2.4 (1.1)**
p [b]	–							*< 0.001*	*0*			*0.006*
PreS	4.6 (1.0)	2.6 (0.7)	3.2 (0.5)	3.8 (1.3)	**4.1 (0.6)**	1.2 (0.6)	0.4 (0.1)	**1.5 (0.3)**	**1.0 (0.5)**	0.3 (0.3)	0	**3.6 (0.9)**
p [b]	–							*< 0.001*	*0.007*			*0.6*
P	3.1 (0.4)	2.1 (0.3)	3.0 (0.3)	1.6 (0.4)	**3.1 (0.3)**	0.8 (0.2)	0.9 (0.1)	**1.3 (0.2)**	**0.3 (0.2)**	0.5 (0.2)	0.1 (0.1)	**2.0 (0.3)**
p [b]	–							*< 0.001*	*0.003*			*0.7*

IG: intragenotypic distances; ISG: intrasubgenotypic distances.
[a] One A and one G isolate were excluded from the PreS and P gene analysis because they were recombinant strains in these regions.
[b] *p*: comparison between D and all the other genotypes; Mann–Whitney's non-parametric test.

by a high frequency of mutants at amino acid level, with a variant S protein being found in more than 80% of cases. In particular, 58% of the patients infected with this genotype carried a virus mutated in the *a determinant*, whereas the prevalence of *a determinant* variants was about ten times less in genotype A, and no substitution was found in genotype G. Likewise, the frequency of *a determinant* mutants in D3 strains was 55%, as against 7.4% in A2 and no mutants in F1.

The *a determinant*, which spans amino acids 124–147 of the surface antigen, is a conformational epitope characterised by two hydrophilic loops that has a common structure in all genotypes (Zuckerman, 2000). It induces neutralising antibodies, and variations in the sequence of the *a determinant* may permit immune evasion by significantly changing its antigenic structure. In particular, mutations in the second loop of the *a determinant* have been frequently described in association with anti-hepatitis B vaccine-induced escape mutants (Zuckerman and Zuckerman, 2003).

The majority of the mutations observed in our isolates were located in the first loop of the *a determinant*, but more than 30% of patients carried strains mutated in the second loop. In particular, two subjects carrying subgenotype D3 and never vaccinated against HBV had strains with a mutation at amino acid 145, which is the prototype "escape mutant" following vaccine-induced immunisation (Carman et al., 1990; Seddigh-Tonekaboni et al., 2001). This finding suggests that escape mutants may be directly acquired as infecting strains or spontaneously arise in hosts under selection forces other than a vaccine-induced immune response; in any case, this finding seems to be epidemiologically relevant and should be considered when designing prevention campaigns.

Eighteen subjects carrying HBV genotype D were also coinfected with HIV-1: there was a particularly high polymorphism of the S and *a determinant* in the isolates obtained from these individuals, although the greater heterogeneity of the D strain was not associated with HIV-1 positivity. D genotype is highly prevalent among Italian IVDUs (Zehender et al., 2003). For this reason one-half of our HIV-positive individuals with the D genotype were IVDUs, and the great majority of them (about 80%), plus two other non-IVDUs, were infected by viral strains characterised by 3–4 mutations in the PreS and *a determinant* region. One of these mutations, T125M, has previously been described by other authors as being associated with IVDUs (Norder et al., 2004). This explains the high degree of polymorphism observed in HBV-D infecting the HIV-1-positive individuals, and suggests the circulation of a specific clade in a particular group of individuals at high risk of infection, possibly representing what remains of the hepatitis B epidemic observed in IVDUs during the late 1970s and 1980s.

One of the possible driving forces of genetic variability is positive selection, which is usually measured by calculating the non-synonymous/synonymous substitution rate ratio over the whole gene sequence. However, as selection generally acts on only one or a few protein sites, maximum likelihood-based methods have recently been developed to evaluate selective pressure at individual sites. ML analysis revealed positively selected sites in our genotype D and (but only in the P coding region) genotype A isolates whereas, albeit on a relatively small number of analysed sequences, the F and G isolates did not show any codons under significant positive selection.

Table 3
Mean synonymous (d_S [a]) and non-synonymous (d_N [a]) (per 10^2 sites) distances of the S, *a determinant*, PreS regions and P by genotype

Region		Genotype							
		D		A		G		F	
		d_S	d_N	d_S	d_N	d_S	d_N	d_S	d_N
S	**Mean**	**4.2**	**2.0**	**1.9**	**0.6**	**0**	**0**	**3.5**	**0.5**
	SE	0.8	0.3	0.5	0.1	0	0	1.1	0.3
a determinant	**Mean**	**7.5**	**2.9**	**3.1**	**0.1**	**0**	**0**	**8.5**	**0.5**
	SE	2.4	0.8	1.2	0.1	0	0	4.0	0.5
PreS	**Mean**	**9.7**	**2.3**	**1.9**	**1.0**	**3.0**	**0.3**	**10.9**	**1.3**
	SE	1.7	0.5	0.6	0.3	1.8	0.3	3.5	0.6
P	**Mean**	**4.3**	**2.6**	**1.8**	**1.1**	**0**	**0.4**	**3.2**	**1.4**
	SE	0.9	0.3	0.4	0.2	0	0.2	0.9	0.3

[a] Estimated using the Nei–Gojobori model. SE: standard error.

Table 4
Frequency of patients carrying mutated strains according to the different HBV genotypes and subgenotypes

Protein	% Variant frequency[a] (mutants/total isolates)											
	D					A			G	F		
	D1 (5)	D2 (8)	D3 (40)	D4 (2)	**Total (55)**	A1 (3)	A2 (27)	**Total (30)**[b]	**(4)**[b]	F1 (3)	F2 (4)	**Total (7)**
S	60 (3)	62.5 (5)	97.5 (39)	100 (2)	**89 (49)**	100 (3)	77.8 (21)	**80 (24)**	**0 (0)**	33.3 (1)	100 (4)	**71.5 (5)**
p[c]	–							0.2	<0.001			0.2
a determinant	40 (2)	100 (8)	55 (22)	0 (0)	**58.2 (32)**	0 (0)	7.4 (2)	**6.7 (2)**	**0 (0)**	0 (0)	100 (4)	**57.2 (4)**
p[c]	–							<0.001	0.02			0.9
PreS	100 (5)	37.5 (3)	75 (30)	100 (2)	**72.7 (40)**	100 (3)	34.6 (9)	**41.4 (12)**	**33.3 (1)**	0 (0)	0 (0)	**0 (0)**
p[c]	–							0.005	0.1			<0.001
P	100 (5)	100 (8)	97.5 (39)	100 (2)	**98.2 (54)**	100 (3)	84.6 (22)	**86.2 (25)**	**33.3 (1)**	100 (3)	100 (4)	**100 (7)**
p[c]	–							0.04	<0.001			0.7

[a] Frequency of patients carrying mutant strains.
[b] One A and one G isolate were excluded from the PreS and P gene analysis because they were recombinant strains in these regions.
[c] p: comparison between genotype D and all the others – χ^2 test.

Genotype D HBV showed positive selective pressure in the S coding region. Interestingly, only one site (at codon 134) under selection was found in the *a determinant*, which frequently showed the substitution of a residue of tyrosine by serine or phenylalanine or (in one case) asparagine all of the other four sites were concentrated in the highly hydrophobic C-terminal of the small S protein, a transmembrane domain of the HBV envelope proteins (Bruss, 2004).

At least in genotype D, mutations in the *a determinant* and C-terminal of the S protein may play a role in escaping cross-immunity, which is one of the main determinants of virus phylogeny (Grenfell et al., 2004). The way in which cross-immunity affected the selection of the mutants at the above sites has yet to be clarified. It has been suggested that the mutation at codon 134 of the *a determinant* described by other authors may be an escape mutant (Ogura et al., 1999; Seddigh-Tonekaboni et al., 2000), and it can be postulated that the mutation in this site (observed in about 13% of our patients carrying genotype D) is driven by the degree of humoral immunity in the population.

The other sites under selection detected in this study were located in a transmembrane domain of the S protein, a region that includes potential T-cell epitopes as suggested for codon 207 (Rehermann et al., 1995), which we found to be under significant positive selection. Mutations at different CTL epitopes have been previously described, but their role in evading CTL responses is still unclear (Rehermann and Nascimbeni, 2005). It can be hypothesised that, in chronically HBV-infected patients, selection pressure at population level in the C-terminal portion of the S protein is attributable to the role that such mutated strains might play in maintaining persistent infection: in particular, the site 207 mutation was found in more than one-third of the genotype D isolates.

Analysis of the P protein revealed several sites under selective pressure, particularly in the rt domain. Two of the three genotype A sites under selective pressure (rt204 and rt180) were associated with lamivudine resistance and, as expected, when the analysis was limited to our lamivudine-treated patients significant selective pressure on these sites was also detected in genotype D polymerase. The frequency of resistant strains seems to be particularly high among patients infected by genotype A (>80%), according with previous studies which suggested that there is a more rapid onset of mutations in its YMDD motif (Buti et al., 2002). It must be born in mind that all of the lamivudine-treated patients in our study population were HIV-1-positive subjects chronically treated with the RT inhibitor at the doses required for antiretroviral therapy, whereas none of the HIV-1-negative subjects was being treated with lamivudine at the time of enrolment.

The relatively small number of sites under selection in the P and S genes may be due to their overlapping, which means that the mutations in both are subject to multiple driving forces (constrained evolution; Mizokami et al., 1997). We found that one site was under significant positive selection at the level of both S (sites 206/207) and P proteins (site rt215), and our analysis suggested that the selective pressure in rt215 is less significant than that in 206 and 207, and is probably a consequence of it. These results emphasise the need for caution when interpreting data obtained in a system in which a mutation in one ORF has consequences in the overlapping ORF (Torresi, 2002). In particular, we found two contiguous sites in the P gene under significant selective pressure in genotypes A (codon rt53) and D (codon rt54), both overlapping codon 45 of the surface antigen, which was highly variable in our case file.

We also attempted to estimate the number and distribution of negatively selected sites, and found that the highest proportion under purifying selection in surface antigen was in the PreS1 and PreS2 proteins, possibly because of their important functional role in cell attachment (Cooper et al., 2003). Nevertheless, deletions in this region (particularly in the PreS2 peptide) were highly prevalent in our patients (6.1%, mainly in genotype D), and it is worth noting that deleted strains have been associated with progression to more severe hepatitis B and the development of liver cancer (Huy et al., 2003; Sugauchi et al., 2003).

There were also two recombinant strains in this region (one A/G and one D/A switch), which confirms other previously reported findings (Bollyky et al., 1996; Kato et al., 2002a; Simmonds and Midgley, 2005) and offers some suggestions for further studies of the evolutionary mechanisms of HBV.

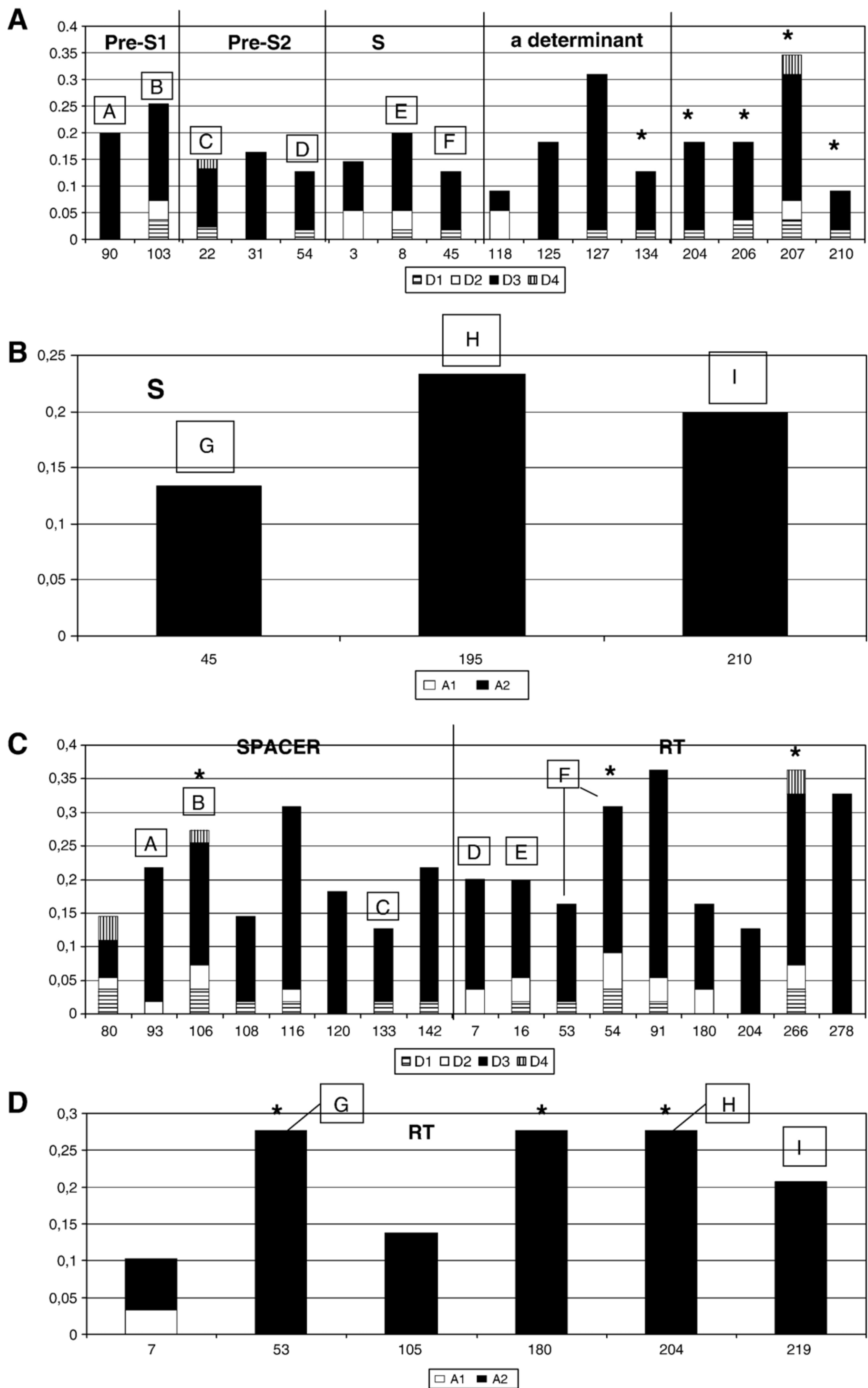

Fig. 3. Frequency of amino acid mutations in PreS/S protein of (A) genotype D and (B) genotype A and in P protein of (C) genotype D and (D) genotype A isolates. Only mutations present in ≥ 10% of the isolates of the same genotype have been represented. X-axis: amino acid sites. Letters indicate S/P overlapping sites. Asterisks indicate sites under selective pressure.

Table 5
Likelihood values, the proportion of sites under selective pressure, and ω values of the S and P coding regions, by viral genotype

Protein	Genotype [a]	ω [b]	LnL [c] of M2 model (p)	ω_2 [b]
S	D	0.74	2846.10 (<0.001)	6.0
	A	0.58	1300.60 (0.05)	5.8
	F	0.35	1053.57 (NS [d])	–
P	D	0.55	5714.04 (<0.001)	3.5
	A	0.64	2449.25 (<0.001)	5.5
	F	0.46	2107.84 (NS)	–

[a] Genotype G was not included because it did not have significant amino acid variations.
[b] ω=mean d_N/d_S for the entire protein; ω_2=mean d_N/d_S for the sites under selection as estimated by ML.
[c] Natural logarithm of the likelihood.
[d] NS: not significant.

In conclusion, genotype D and its subgenotypes are more heterogeneous than the other genotypes/subgenotypes. At least part of this heterogeneity seems to be due to selection at population level, but it is likely to be driven by mechanisms other than selection, including greater replication activity, longer evolution and, possibly, wider dispersion in the general population.

Our finding that more patients carry D strains with mutations in the *a determinant* of antigen S, and the isolation of vaccine-induced prototypic “escape” mutants in non-vaccinated subjects, indicate the need for careful surveillance of these variants in areas in which genotype D predominates.

Finally, a number of studies have suggested that the outcome of the hepatitis due to genotype D is more severe than that due to genotype A (Sanchez-Tapias et al., 2002; Thakur et al., 2002; Wai et al., 2005), and that genotype D is less responsive to alpha-interferon (Erhardt et al., 2005). Further studies are needed to clarify the possible relationship between the heterogeneity and greater virulence of this genotype.

Materials and methods

Patient characteristics

The study involved 99 hepatitis B surface antigen (HBsAg)-positive patients (73 males and 26 females) with chronic HBV infection and detectable serum HBV–DNA, who were living in Northern Italy and were observed in the participating hospital units between 1993 and 2005. All of the patients had had high serum alanine aminotransferase (ALT) levels for more than 6 months at the time of enrolment. Six patients were born outside Italy: three in Asia (China, Japan and Indonesia), and one each in Brazil, Nigeria and Great Britain.

HBV–DNA extraction and amplification by nested PCR

Viral DNA was extracted from the patients' sera (stored at −30 °C) using the High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH, Mannheim, Germany) as recommended by the manufacturer. The DNA was eluted in 50 μL of nuclease-free distilled water. About 10 μL of the extracted DNA were amplified in a 50 μL mixture of 1× PCR buffer (10 mmol/L Tris–HCl, pH 8.3, 50 mmol/L KCl), 1.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTPs, 0.5 μmol/L of the outer primers, and 1 U of Ampli*Taq* Gold DNA polymerase (Applied Biosystems, Foster City, CA). Second-round PCR was performed using 5 μL of the first round PCR product in 50 μL of the same mixture containing 0.5 μmol/L of the inner primers.

All of the amplification procedures were performed in an automated thermal-cycler (Applied Biosystems) under similar conditions: a long initial DNA denaturation step at 95 °C for 9 min, followed by specific cycling conditions, a prolonged extension step at 72 °C for 10 min, and finally by holding at 4 °C.

Optimal positive and negative controls were included in each extraction and amplification run.

About 10 μL of the nested PCR products mixed with Blue/Orange loading dye 6× (Promega, Madison, WI USA) were run on a ethidium bromide-stained 1.5% agarose gel together with 0.5 μg Φ×174 DNA/Hae III marker (Promega). The expected bands were detected by means of a UV transilluminator.

Primers

A sequence of about 1365 nucleotides (nct, from nct 2474 to nct 1108 of full-length HBV genome) was obtained by means of nested PCR amplification using five sets of degenerated primers to obtain two large outer fragments and three inner and partially overlapping sequences. The primer names, sequences and PCR conditions are shown in Table 1. The nucleotide numeration refers to isolate NDR260 (Okamoto et al., 1988).

HBV–DNA sequencing

The HBV amplicons were separated from the residual dNTPs and primers using a commercial purification kit (QIAquick PCR Purification Kit, Qiagen GmbH, Hilden, Germany). Subsequently, the nucleotide sequences of the three amplified products were determined in both directions by means of the dideoxy terminator method in the presence of the specific sense or anti-sense inner primer, using the BigDye Terminator Cycle Sequencing Ready Reaction Kit and the automated DNA Sequencer 3100 Genetic Analyzer (Applied Biosystems). In brief, an optimal quantity of purified PCR product was mixed with 3.2 pmol of the sequencing primer and 8 μL of the ready reaction mix in a final volume of 20 μL. Cycle sequencing reactions were performed for 25 cycles, each of 20 s at 95 °C, 20 s at 55 °C, and 3 min at 60 °C. Excess dye terminators were removed from the sequencing mixture by means of ethanol/sodium acetate precipitation. Heat-dried samples were dissolved in 10 μL of nuclease-free distilled water and loaded onto the DNA sequencer according to the manufacturer's instructions.

Phylogenetic and statistical analyses

Phylogenetics

The phylogenetic analysis was made on a sequence of 1182 nucleotides encompassing the ORFs coding for the entire S (including *a determinant*: amino acids 110–160)

and PreS2 proteins, and a tract of 59 amino acids at the C-terminal of PreS1, as well as the overlapping ORF coding for a fragment of 394 amino acids of viral polymerase, including 106 residues of C-terminal spacer and amino acids 1 to 288 of rt.

The 99 patient isolates were aligned with reference strains retrieved from the GenBank database and representing all major HBV genotypes and subgenotypes (see below) using the CLUSTALW program (Thompson et al., 1994) integrated within a biological sequence alignment editor (Bio-Edit by Tom Hall, 2001).

The analyses were made using Phylip vs. 3.5 (Felsenstein, 1989) and PAUP* v. 4.0b10 (by Swofford D., 2001, Sinauer Ass. Inc. Publ.) software packages.

Best-fitting substitution models and parameters were estimated using a hierarchical likelihood ratio test implemented in the Modeltest 3.07 computer program (Posada and Crandall, 1998), selecting the transversion model (TVM) that considers different transversion and equal transition rates, plus invariable sites (proportion = 0.49) and γ-distributed rates among sites (α-shape parameter = 0.71). Distances were expressed in terms of the number of substitutions per site (or 10^2 sites when indicated). The phenetic analysis was made by grouping the genetic distances by isolate genotype or subgenotype (intragenotypic – IG or intrasubgenotypic – ISG mean distances).

The phylogeny was reconstructed using the neighbor-joining (NJ) and maximum-likelihood (ML) methods, and the topologies confirmed by the two methods were considered significant. The significances of the interior branches in the NJ were tested by means of bootstrap analysis with 1000 replicates; bootstrap values of ≥70% were considered significant. The significance of the branch length in ML analysis was estimated using the likelihood ratio test for a branch length of zero, implemented in PAUP* (where $0.01 < p < 0.05$ is significant, and $p < 0.01$ is highly significant).

The deduced amino acid sequences were aligned with reference strains by subgenotype, and the majority-rule consensus sequence was calculated. Each isolate that differed from the consensus sequence was defined a mutant.

The inconsistencies in numbering the amino acid sites are due to the genome-length polymorphism of HBV genotypes. In order to avoid ambiguities, we restarted numbering PreS/S proteins at the first codon position of each domain. We used the new nomenclature proposed by Stuyver for P protein (Stuyver et al., 2001).

Synonymous (d_S) and nonsynonymous (d_N) distances were estimated using the model of Nei and Gojobori (1986) with the Jukes–Cantor correction and MEGA version 3.1 (Kumar et al., 2004).

Site-specific selection was evaluated using the following models implemented in the CODEML program included in the PAML v.3.14 package (Yang, 1997), which allow for heterogeneous $\omega = d_N/d_S$ ratios among codons: M0 (one-ratio), assuming one d_N/d_S ratio for all sites; M1 (nearly-neutral), assuming a proportion of conserved and a proportion of neutral sites ($0 < \omega < 1$ or $\omega = 1$); and M2 (selection), adding an extra class of sites under positive selection ($\omega > 1$) and representing a slight modification of the previously described M1 and M2 models (Nielsen and Yang, 1998). The performances of the models were compared using the likelihood-ratio test, which allows a comparison of the likelihood of nested models. An F3×4 codon frequency model was assumed, and k was estimated by the program. A new empirical Bayes approach (Bayes Empirical Bayes – BEB) was used to identify specific sites under positive selection, with a posterior probability of ≥95% (Yang et al., 2005). Isolates showing genotypic recombinations, nucleotide deletions or premature stop codons were excluded from the analysis.

Negative selection was evaluated by means of single-likelihood ancestor counting (SLAC) analysis, which is based on the method originally described by Suzuki (Suzuki and Gojobori, 1999) and implemented in the HYPHY software package (Kosakovsky Pond and Frost, 2005; Pond et al., 2005).

Strains used for the phylogenetic analysis. The isolates were aligned with 74 reference strains representative of the major HBV genotypes (A–G) and the subgenotypes most frequently found in our study population (A1–3; D1–5; F1,2), whose accession numbers are: X70185 (Preisler-Adams et al., 1993b); X51970; M54923; X02763, D00329, AB033558, AB033559 (Okamoto et al., 1988); X01587 (Fujiyama et al., 1983); X02496 (Bichko et al., 1985); X65257; X65259; AY230128; X75657, X75658 (Norder et al., 1994); AB036907, AB036910, AB036912, AB036919, AB036920 (Nakano et al., 2001); AF160501 (Stuyver et al., 2000); AF405706; AB064310, AB064311, AB064312, AB064313 (Kato et al., 2002b); AB056513, AB056515 (Kato et al., 2001); AB048701, AB048703 (Sugauchi et al., 2001); X80925; Z35716; AB194950, AB194949, AB194951, AB194952(Kurbanov et al., 2005);AY090453, AY090458, AY090456, AY090459, AY090255 (Arauz-Ruiz et al., 2002); V00866 (Ono et al., 1983); Y07587 (Stoll-Becker et al., 1997); AB078032, AB078033 (Chen et al., 2003); AF143298 (Preikschat et al., 1999); M32138 (Tong et al., 1990); V01460 (Galibert et al., 1979); AF151735 (Gerner et al., 1999); S50225 (Wands et al., 1992); X69798 (Naumann et al., 1993); AB090270, AB090269, AB090268 (Duong et al., 2004); AF297621 (Owiredu et al., 2001); U87742 (Bowyer et al., 1997); X72702 (Preisler-Adams et al., 1993a); AB064314; AB086397; AF223963; AF223962; AF223965; AF280817; AF418674; AF418682; AJ309371; AY161150; AY161161; AY161157; AY161140; X85254; X68292; Z72478; M57663; AJ344116.

Sequence accession numbers: the sequences were submitted to GenBank with assigned accession numbers from EF514251 to EF514349 and EF522104.

Statistical analysis

Mean genetic distances were compared using Mann–Whitney's non-parametric test for independent samples. p values of ≤0.05 were considered significant. All of the analyses were performed using the SPSS package v. 13 (SPSS Inc., Chicago, IL, USA) or epi-info v.6 (CDC, Atlanta).

Acknowledgments

The authors would like to thank the Italian Association against AIDS (ANLAIDS–Sezione Lombarda) for its continuous support, and Dr. Paolo Sacchi (Division of Infectious and Tropical Diseases, IRCCS San Matteo Hospital, Pavia) for collaborating with the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2007.03.015.

References

Arauz-Ruiz, P., et al., 1997. Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variability of the small S gene. J. Infect. Dis. 176, 851–858.

Arauz-Ruiz, P., et al., 2002. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J. Gen. Virol. 83, 2059–2073.

Banerjee, A., et al., 2006. Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. J. Med. Virol. 78, 1164–1174.

Bichko, V., et al., 1985. Subtype ayw variant of hepatitis B virus. DNA primary structure analysis. FEBS Lett. 185, 208–212.

Bollyky, P.L., et al., 1996. Recombination between sequences of hepatitis B virus from different genotypes. J. Mol. Evol. 42, 97–102.

Bowyer, S.M., et al., 1997. A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. J. Gen. Virol. 78, 1719–1729.

Bruss, V., 2004. Envelopment of the hepatitis B virus nucleocapsid. Virus Res. 106, 199–209.

Buti, M., et al., 2002. Is hepatitis B virus subtype testing useful in predicting virological response and resistance to lamivudine? J. Hepatol. 36, 445–446.

Carman, W.F., et al., 1990. Vaccine-induced escape mutant of hepatitis B virus. Lancet 336, 325–329.

Chattopadhyay, S., et al., 2006. Hepatitis B virus genotypes in chronic liver disease patients from New Delhi, India. World J. Gastroenterol. 12, 6702–6706.

Chen, Y., et al., 2003. Complete genome sequence of hepatitis B virus (HBV) from a patient with fulminant hepatitis without precore and core promoter mutations: comparison with HBV from a patient with acute hepatitis infected from the same infectious source. J. Hepatol. 38, 84–90.

Cooper, A., et al., 2003. The earliest steps in hepatitis B virus infection. Biochim. Biophys. Acta 1614, 89–96.

Duong, T.N., et al., 2004. Comparison of genotypes C and D of the hepatitis B virus in Japan: a clinical and molecular biological study. J. Med. Virol. 72, 551–557.

Erhardt, A., et al., 2005. Response to interferon alpha is hepatitis B virus genotype dependent: genotype A is more sensitive to interferon than genotype D. Gut 54, 1009–1013.

Felsenstein, J., 1989. Phylogeny inference package (version 3.2). Cladistics 5, 164–166.

Fujiyama, A., et al., 1983. Cloning and structural analyses of hepatitis B virus DNAs, subtype adr. Nucleic Acids Res. 11, 4601–4610.

Galibert, F., et al., 1979. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. Nature 281, 646–650.

Gerner, P., et al., 1999. Hepatitis B virus core promoter mutations in children with multiple anti-HBe/HBeAg reactivations result in enhanced promoter activity. J. Med. Virol. 59, 415–423.

Grenfell, B.T., et al., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332.

Huy, T.T., et al., 2003. High prevalence of hepatitis B virus pre-s mutant in countries where it is endemic and its relationship with genotype and chronicity. J. Clin. Microbiol. 41, 5449–5455.

Kao, J.H., et al., 2000. Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. Gastroenterology 118, 554–559.

Kato, H., et al., 2001. Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. J. Virol. Methods 98, 153–159.

Kato, H., et al., 2002a. Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. Hepatology 35, 922–929.

Kato, H., et al., 2002b. Characteristics of hepatitis B virus isolates of genotype G and their phylogenetic differences from the other six genotypes (A through F). Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. J. Virol. 76, 6131–6137.

Kidd-Ljunggren, K., et al., 2002. Genetic variability in hepatitis B viruses. J. Gen. Virol. 83, 1267–1280.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.

Kramvis, A., et al., 2002. Analysis of the complete genome of subgroup A′ hepatitis B virus isolates from South Africa. J. Gen. Virol. 83, 835–839.

Kumar, S., et al., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Kumar, A., et al., 2005. Hepatitis B virus genotype A is more often associated with severe liver disease in northern India than is genotype D. Indian J. Gastroenterol. 24, 19–22.

Kurbanov, F., et al., 2005. A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. J. Gen. Virol. 86, 2047–2056.

Li, J.S., et al., 1993. Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. J. Virol. 67, 5402–5410.

Makuwa, M., et al., 2006. Identification of hepatitis B virus subgenotype A3 in rural Gabon. J. Med. Virol. 78, 1175–1184.

McMillan, J.S., et al., 1996. Mutations in the hepatitis B virus precore/core gene and core promoter in patients with severe recurrent disease following liver transplantation. Hepatology 24, 1371–1378.

Mizokami, M., et al., 1997. Constrained evolution with respect to gene overlap of hepatitis B virus. J. Mol. Evol. 44, S83–S90.

Nakano, T., et al., 2001. Characterization of hepatitis B virus genotypes among Yucpa Indians in Venezuela. J. Gen. Virol. 82, 359–365.

Naumann, H., et al., 1993. Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J. Gen. Virol. 74, 1627–1632.

Nei, M., Gojobori, T., 1986. Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Norder, H., et al., 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., et al., 2004. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. Intervirology 47, 289–309.

Ogura, Y., et al., 1999. Prevalence and significance of naturally occurring mutations in the surface and polymerase genes of hepatitis B virus. J. Infect. Dis. 180, 1444–1451.

Okamoto, H., et al., 1987. Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. Jpn. J. Exp. Med. 57, 231–236.

Okamoto, H., et al., 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J. Gen. Virol. 69, 2575–2583.

Ono, Y., et al., 1983. The complete nucleotide sequences of the cloned hepatitis B virus DNA; subtype adr and adw. Nucleic Acids Res. 11, 1747–1757.

Orito, E., et al., 1989. Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. Proc. Natl. Acad. Sci. U.S.A. 86, 7059–7062.

Owiredu, W.K., et al., 2001. Molecular analysis of hepatitis B virus genomes isolated from black African patients with fulminant hepatitis B. J. Med. Virol. 65, 485–492.
Pond, S.L.K., et al., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.
Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Preikschat, P., et al., 1999. Hepatitis B virus genomes from long-term immunosuppressed virus carriers are modified by specific mutations in several regions. J. Gen. Virol. 80, 2685–2691.
Preisler-Adams, S., et al., 1993a. Sequence analysis of hepatitis B virus DNA in immunologically negative infection. Arch. Virol. 133, 385–396.
Preisler-Adams, S., et al., 1993b. Complete nucleotide sequence of a hepatitis B virus, subtype adw2, and identification of three types of C open reading frame. Nucleic Acids Res. 21, 2258.
Rehermann, B., Nascimbeni, M., 2005. Immunology of hepatitis B virus and hepatitis C virus infection. Nat. Rev., Immunol. 5, 215–229.
Rehermann, B., et al., 1995. Hepatitis B virus (HBV) sequence variation of cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. J. Clin. Invest. 96, 1527–1534.
Rodriguez-Frias, F., et al., 1995. Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. Hepatology 22, 1641–1647.
Sanchez-Tapias, J.M., et al., 2002. Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in western patients. Gastroenterology 123, 1848–1856.
Schaefer, S., 2005. Hepatitis B virus: significance of genotypes. J. Viral Hepatitis 12, 111–124.
Seddigh-Tonekaboni, S., et al., 2000. Effect of variation in the common "a" determinant on the antigenicity of hepatitis B surface antigen. J. Med. Virol. 60, 113–121.
Seddigh-Tonekaboni, S., et al., 2001. Hepatitis B surface antigen variants in vaccines, blood donors and an interferon-treated patient. Effect of variation in the common "a" determinant on the antigenicity of hepatitis B surface antigen. J. Viral Hepatitis 8, 154–158.
Shiina, S., et al., 1991a. Relationship of HBsAg subtypes with HBeAg/anti-HBe status and chronic liver disease: Part II. Evaluation of epidemiological factors and suspected risk factors of liver dysfunction. Relationship of HBsAg subtypes with HBeAg/anti-HBe status and chronic liver disease. Part I: analysis of 1744 HBsAg carriers. Am. J. Gastroenterol. 86, 872–875.
Shiina, S., et al., 1991b. Relationship of HBsAg subtypes with HBeAg/anti-HBe status and chronic liver disease: Part I. Analysis of 1744 HBsAg carriers. Am. J. Gastroenterol. 86, 866–871.
Simmonds, P., Midgley, S., 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. 79, 15467–15476.
Stoll-Becker, S., et al., 1997. Transcription of hepatitis B virus in peripheral blood mononuclear cells from persistently infected patients. J. Virol. 71, 5399–5407.
Stuyver, L., et al., 2000. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J. Gen. Virol. 81, 67–74.
Stuyver, L.J., et al., 2001. Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. Hepatology 33, 751–757.
Sugauchi, F., et al., 2001. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. J. Gen. Virol. 82, 883–892.
Sugauchi, F., et al., 2003. Influence of hepatitis B virus genotypes on the development of preS deletions and advanced liver disease. J. Med. Virol. 70, 537–544.
Sumi, H., et al., 2003. Influence of hepatitis B virus genotypes on the progression of chronic type B liver disease. Hepatology 37, 19–26.
Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16, 1315–1328.
Thakur, V., et al., 2002. Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in the Indian subcontinent. J. Gastroenterol. Hepatol. 17, 165–170.
Thompson, J.D., et al., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Tong, S.P., et al., 1990. Active hepatitis B virus replication in the presence of anti-HBe is associated with viral variants containing an inactive pre-C region. Virology 176, 596–603.
Torresi, J., 2002. The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. J. Clin. Virol. 25, 97–106.
Tsubota, A., et al., 2001. Genotype may correlate with liver carcinogenesis and tumor characteristics in cirrhotic patients infected with hepatitis B virus subtype adw. J. Med. Virol. 65, 257–265.
Wai, C.T., et al., 2005. Clinical outcome and virological characteristics of hepatitis B-related acute liver failure in the United States. J. Viral Hepatitis 12, 192–198.
Wands, J.R., et al., 1992. Molecular pathogenesis of liver disease during persistent hepatitis B virus infection. Semin. Liver Dis. 12, 252–264.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.
Yang, Z., et al., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118 (Electronic publication 2005 Feb 2).
Yuen, M.F., et al., 2003. Significance of hepatitis B genotype in acute exacerbation, HBeAg seroconversion, cirrhosis-related complications, and hepatocellular carcinoma. Hepatology 37, 562–567.
Zehender, G., et al., 2003. Hepatitis B virus genotype distribution in HIV-1 coinfected patients. Gastroenterology 125, 1559–6015 (author reply 1660).
Zollner, B., et al., 2002. Subtype-dependent response of hepatitis B virus during the early phase of lamivudine treatment. Clin. Infect. Dis. 34, 1273–1277 (Electronic publication 2002 Apr 2).
Zuckerman, A.J., 2000. Effect of hepatitis B virus mutants on efficacy of vaccination. Lancet 355, 1382–1384.
Zuckerman, J.N., Zuckerman, A.J., 2003. Mutations of the surface protein of hepatitis B virus. Antiviral Res. 60, 75–78.