

The prevalence of mutations in the major hydrophilic region of the surface antigen of hepatitis B virus varies with subgenotype

X. Y. WANG¹, T. J. HARRISON², X. HE³, Q. Y. CHEN¹, G. J. LI⁴, M. H. LIU^{1,5},
H. LI¹, J. Y. YANG¹ AND Z. L. FANG^{1*}

¹ Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control, Nanning, Guangxi, China

² Division of Medicine, UCL Medical School, London, UK

³ Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China

⁴ Department of Public Health of Guangxi Zhuang Autonomous Region, Nanning, Guangxi, China

⁵ School of Preclinical Medicine, Guangxi Medical University, Nanning, Guangxi, China

Received 27 November 2014; Final revision 16 January 2015; Accepted 26 January 2015;
first published online 23 April 2015

SUMMARY

Mutations in the major hydrophilic region (MHR) of the surface antigen of hepatitis B virus (HBV) may result in vaccine escape, failure of immunotherapy and antiviral resistance. These mutants may be transmitted and constitute a public health threat. We aimed to determine the prevalence of MHR mutations of HBV in areas of high endemicity in Guangxi, China. HBV surface gene was analysed from 278 HBsAg-positive asymptomatic individuals recruited from Guangxi using cluster sampling. Three genotypes, B, C and I, were identified. The overall prevalence of MHR mutations is 17.6%. The prevalence of MHR mutations in genotype B (15.1%) is not significantly different from that in genotype C (16.4%). However, the prevalence in subgenotype C5 (31.1%) is significantly higher than in subgenotype C2 (13.0%) ($\chi^2 = 6.997$, $P < 0.05$). The prevalence of escape mutations and overlapping polymerase substitutions in subgenotype C5 is significantly higher than in subgenotypes B2 and C2. In total, 7.9% of MHR mutants are escape mutations and 72.1% of MHR mutations produced amino-acid changes in the overlapping polymerase, including resistance mutations to entecavir. Our results suggest that the prevalence of MHR mutations varies with subgenotype. The prevalence of escape mutations and polymerase mutations may be associated with subgenotype.

Key words: Hepatitis B virus, major hydrophilic region, mutations, prevalence, subgenotypes.

INTRODUCTION

HBV has a circular, partially double-stranded DNA genome of about 3200 nt with four open reading frames (ORFs), namely the core/precore, polymerase,

surface and X ORF [1]. The longest ORF encodes the viral polymerase (Pol; Pol-ORF), which encodes a reverse transcriptase (RT) activity that is essential for replication and is the target of nucleos(t)ide analogues (NA). The surface ORF is located within the Pol-ORF but in a different reading frame. This encodes the pre-S1, pre-S2 and S proteins (designated large, middle and small HBsAg, respectively) [2]. These three proteins share 226 carboxyl-terminal amino-acid residues. The central region of S-HBsAg

* Author for correspondence: Dr Z. L. Fang, Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control, 18 Jin Zhou Road, Nanning, Guangxi, China, 530028.
(Email: zhongliaofang@hotmail.com)

(amino acids 99–169) is the major hydrophilic region (MHR) exposed on the surfaces of virions and sub-viral particles and harbours many B-cell epitopes [3].

The MHR contains the 'a' determinant (residues 124–147) which was defined originally as the antigenic region shared by all serological variants of HBV and is the region primarily associated with the induction of a protective humoral immune response [4] and the immunodominant region of HBsAg [5]. Amino-acid substitutions and multiple changes in the 'a' determinant can modify the antigenicity and immunogenicity of hepatitis B virus (HBV) resulting in failure to react in immunoassays and antibody escape [6, 7]. Furthermore, the susceptibility of these variants to antiviral therapy may be altered because mutations in the surface ORF may result in amino-acid substitutions in the polymerase [8].

The lack of a proofreading activity of the viral polymerase leads to a high rate of mutation during replication of the HBV genome. Traditionally, MHR mutants have been thought to be selected by the pressure of hepatitis B immunoglobulin, immunization or antiviral therapy. However, these variants also may occur naturally [9]; they are stable and can be transmitted horizontally and vertically [10] and their circulation in the population may represent a public health threat [11].

Guangxi was one of the first provinces in China to introduce universal immunization of all newborn infants with hepatitis B vaccine. In 2009, the Guangxi Government also carried out a free, mass catch-up immunization programme for those children aged <15 years who had not been immunized at birth. Although the prevalence of HBsAg in the entire general population of China was 7.2% in 2006, that of Guangxi was 9.2% in 2011 [12]. Clearly, despite the mass immunization programme, Guangxi remains in the category of high endemicity. However, data regarding the prevalence of MHR variants and subgenotypes is lacking.

In this cross-sectional study, we aimed to determine the prevalence of MHR variants and subgenotypes of HBV in the general population throughout Guangxi.

MATERIALS AND METHODS

Study subjects and sample design

The study subjects were selected from a study cohort, which has been described previously [12]. Briefly, the cohort was established by recruiting study subjects from a rural area of Guangxi province using stratified

random cluster sampling, then a traditional village with a population of about 1000 was selected randomly from each county and inhabitants of all ages were recruited. All study subjects in the cohort whose stored serum samples remained available were included in the analysis. Informed, written consent was obtained from each individual. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Guangxi Institutional Review Board.

Serological testing

Serological testing was conducted in Guangxi Centre for Disease Prevention and Control. Sera were tested for HBsAg/anti-HBs, HBeAg/anti-HBe and anti-HBc using enzyme immunoassays (Beijing Wantai Biological Pharmacy Enterprise, China).

Nested PCR for HBV DNA and nucleotide sequencing

DNA was extracted from 85 μ l serum by pronase digestion followed by phenol/chloroform extraction. First-round polymerase chain reaction (PCR) was carried out in a 50 μ l reaction using primers LSOB1 (nt 2739–2762, 5'-GGCATTATTTGCATACCCTTTGG-3') and P2 (nt 1823–1806, 5'-CCGAAAGCTTGAGCTCTTCAAAAAGTTGCATGGTGCTGG-3') [13], with 5 min hot start followed by 30 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 90 s. Second-round PCR was carried out on 5 μ l of the first-round products in a 50 μ l reaction using primers LSB11 (nt 2809–2829, 5'-TTGTGGGTCACCATATTCTT-3') and POLSEQ3 (nt 1358–1339, 5'-CCGCGAGAGGACGACAGAAT-3') with the same amplification protocol as first round. Products from the second round were confirmed by agarose gel electrophoresis. HBV DNA-positive products were sent to Sangon Biotech (China) for sequencing using a BigDye Terminator v. 3.1 Cycle Sequencing kit (Applied Biosystems, USA) with sequencing primers POLSEQ3 and MD14 (nt 418–433, 5'-GCGCTGCAGCTATGCCTCATCTTC-3').

HBV genotyping

HBV genotypes were determined using phylogenies reconstructed on the basis of the complete S region (678 nt) of the viruses. The sequences were aligned to 56 HBV sequences of all known genotypes retrieved from GenBank using Clustal W, and visually

confirmed with the sequence editor BioEdit [14]. These sequences include B6-AB287314-USA, B6-AB287320-Greenland, B7-GQ358138-Indonesia, B8-AP011096-Indonesia, B9-GQ358152-Indonesia, C6-AB493844-Indonesia, C6-AB493843-Indonesia, C8-AP011107-Indonesia, C10-AB540583-Indonesia, C11-AB554020-Indonesia, C12-ABAB554018-Indonesia, C13-AB644281-Indonesia, C14-AB644283-Indonesia, AY226578-Woolly monkey, A1-M57663-Philippines, B9-GQ358149-Indonesia, B8-GQ358147-Indonesia, B7-GQ358137-Indonesia, B1-D23677-Japan, B2-AY217358-China, B2-AF121249-Vietnam, B3-AB033555-Sumatra, B4-AB073835-Vietnam, B5-AB219427-Philippines, C1-AF458664-China, C2-AY217371-China, C3-X75656-Polynesia, C4-AB048704-Australia, C5-AB241109-Philippine, B1-AB073845-Japan, B3-M54923-Sulawesi, B4-AB100695-Vietnam, B5-AB241117-Philippines, C1-AB050018-Japan, C2-AF223960-Malaysia, C3-X75665-New Caledonia, C4-AB048705-Australia, C5-AB241111-Philippine, C16-AB644287-Indonesia, C15-AB644286-Indonesia, C14-AB644284-Indonesia, C13-AB644280-Indonesia, C12-AB554025-Indonesia, C11-AB554019-Indonesia, C9-AP011108-Indonesia, C8-AP011106-Indonesia, C7-GU721029-South Korea, C7-EU670263-Philippines, D1-Y07587-Germany, E-AB032431-Liberia, F-AY090456-Nicaragua, G-AF160501-USA, H-AY090460-USA, I1-FJ023659-Laos, I2-FJ023664-Laos, J-AB486012-Japan. Neighbour-joining trees were reconstructed under the Kimura two-parameter substitution model with the MEGA program [15]. The reliability of clusters was evaluated using an interior branch test with 1000 replicates, and the internal nodes with over 95% support were considered reliable.

Identification of MHR mutations and overlapping RT mutations

MHR mutations were originally evaluated using the Genafor/Arevir-geno2pheno drug resistance tool. Those identified as MHR mutations by the tool were then aligned to HBV reference sequences (AB073835 and AF223960), which were obtained from GenBank and used to exclude subgenotypes and polymorphisms. The mutations were categorized into general mutations and escape mutations.

Statistical analysis

Statistical comparisons of the prevalence of MHR mutations, between men and women and across the various

genotypes and different counties were performed using χ^2 tests. The 95% confidence intervals (CIs) for the prevalence of MHR mutations and genotypes were estimated. All *P* values were two-tailed and *P* < 0.05 was considered to be significant. Logistic regression analysis was performed to determine factors associated with MHR mutations. Statistical analysis was performed using Epi Info v. 6.1 (CDC, USA) and SPSS v. 13.0 (SPSS Inc., USA) software.

RESULTS

Baseline characteristics

The study subjects were 278 HBsAg-positive asymptomatic individuals, including 153 males and 125 females. The youngest and oldest were aged 3 and 89 years, respectively, and their average age (mean \pm s.d.) was 34.9 ± 18.7 years. Eighteen individuals were positive for both HBsAg and anti-HBs (two were immunized, 12 were not immunized and the immunization history of four was unknown). Forty-one had been immunized against HBV but 167 had not received the vaccine. The immunization history of the remaining 70 is unclear. None of these individuals had received antiviral or immunoglobulin therapy (Table 1).

Overall prevalence of MHR mutations according to sex and age

The overall prevalence of MHR mutations was 17.6% (49/278, 95% CI 13.1–22.1). The prevalence in males and females was 19.0% (95% CI 12.8–25.2) and 16.0% (95% CI 9.6–22.4), respectively. There was no significant difference between males and females ($\chi^2 = 0.414$, *P* > 0.05). The overall prevalence of escape mutations involved in diagnostic failures or in immunization with hepatitis B vaccine and hepatitis B immunoglobulin (HBIG) therapy within the MHR was 7.9% (22/278, 95% CI 4.7–11.1). The prevalence of escape mutations in males and females was 9.8% (95% CI 5.1–14.5) and 5.6% (95% CI 1.6–9.6), respectively. Again, no significant difference was found between males and females ($\chi^2 = 1.668$, *P* > 0.05), suggesting that MHR mutations are not associated with sex. The prevalence of MHR mutations did not vary with age.

Prevalence of MHR mutations according to genotypes/subgenotypes

Two hundreds and seventy-two of the 278 samples in this study could be genotyped. Three genotypes were

Table 1. Characteristics of the study subjects

	Overall	Subgenotype B2	Subgenotype C2	Subgenotype C5
Total, <i>N</i>	278	70	108	45
Sex, male, <i>n</i> (%)	153 (55.0%)	44 (62.9%)	62 (57.4%)	20 (44.4%)
Age (mean \pm s.d.)	34.9 \pm 18.7	34.0 \pm 20.0	34.0 \pm 16.8	36.1 \pm 20.1
Vaccine				
No, <i>n</i> (%)	167 (60.1%)	38 (22.8%)	66 (39.5%)	29 (17.4%)
Unknown, <i>n</i> (%)	70 (25.2%)	23 (32.9%)	25 (35.7%)	8 (11.4%)
Immunized, <i>n</i> (%)	41 (14.7%)	9 (22.0%)	17 (41.5%)	7 (17.1%)
Abnormal ALT*, % (<i>n/N</i>)	23.9% (50/209)	27.3% (15/55)	29.6% (24/81)	3.3% (1/30)
Anti-HBs positivity, % (<i>n/N</i>)	6.6% (18/272)	5.7% (4/70)	6.8% (7/103)	13.6% (6/44)
HBeAg positivity, % (<i>n/N</i>)	31.6% (86/272)	31.4% (22/70)	43.7% (45/103)	25.0% (11/44)

ALT, Alanine aminotransferase.

* The cut-off value for 'abnormal ALT' is ≥ 40 IU.

found, B, C and I. Eighty-six belonged to genotype B, of which 70 and two were assigned to subgenotypes B2 and B4, respectively, but the remaining 14 could not be subgenotyped. Subgenotypes B2 and B4 account for 81.4% (70/86) and 2.3% (2/86), respectively. One hundred and eighty-three samples belonged to genotype C and subgenotypes C1, C2, C5 and C10 were found in 7, 108, 45 and 3, respectively. Subgenotypes C1, C2, C5 and C10 accounted for 3.8% (7/183), 59.0% (108/183), 24.6% (45/183) and 1.6% (3/183), respectively. The remaining 20 could not be subgenotyped. Three belonged to genotype I (Supplementary Fig. S1a–c).

The prevalence of MHR mutations in genotype B is not significantly different from that in genotype C ($\chi^2 = 0.071$, $P > 0.05$) (Table 2). In the stratification analysis according to subgenotype, there is also no significant difference in the prevalence of MHR mutations between subgenotype B2 and subgenotype C2 ($\chi^2 = 0.266$, $P > 0.05$). However, the prevalence of MHR mutations in subgenotype C5 is significantly higher than in genotype C2 ($\chi^2 = 6.997$, $P < 0.05$), and is also higher than in subgenotype B2, although the difference is not significant ($\chi^2 = 3.817$, $P > 0.05$). It is interesting that there is also no significant difference in the prevalence of escape mutations between subgenotypes B2 and C2 ($\chi^2 = 0.0008$, $P > 0.05$) but the prevalence of escape mutations in subgenotype C5 is significantly higher than in subgenotypes B2 and C2 ($\chi^2 = 5.575$, $P < 0.05$; $\chi^2 = 9.027$, $P < 0.05$, respectively) (Table 2). These data suggest that the prevalence of MHR mutations in the viral surface ORF varies with HBV subgenotype. Subgenotype C5 has a higher prevalence of MHR mutations and escape mutations.

Geographical distribution of the prevalence of MHR mutations

The prevalence of MHR mutations varies between the counties. The highest prevalence is in Napo county (35.4%, 95% CI 21.9–48.9) and the lowest is in Cangwu county (7.4%, 95% CI 4.2–14.4) (Fig. 1). The difference in the prevalence of MHR mutations is significant between Napo and Cangwu counties ($\chi^2 = 12.194$, $P < 0.05$), suggesting that the prevalence of MHR mutations of the viral surface ORF may also vary geographically.

Distribution of the mutations within the MHR

Forty-nine of the 278 study subjects had amino-acid substitutions in the MHR and 21 (42.9%) had amino-acid substitutions involving the 'a' determinant. Of these 49 subjects, 32 (65.3%) had a single mutation and 17 (34.7%) had a combination of 2–7 mutations, making a total of 43 mutations. Fourteen of the 43 mutations were located within the 'a' determinant, including nine in the first loop (amino acids 124–137) and five in the second loop (amino acids 139–147) (Fig. 2). The most frequent substitutions were A159 V and V168A, which were found in 2.2% (6/278). The next frequent substitutions were S114 T, P120 T and Q129H, which were found in 1.4% (4/278).

Variants which have been implicated in escape from passively acquired (via immunotherapy) or vaccine-induced antibody was found in 22 subjects (7.9%, 95% CI 4.7–11.1). Of these 22 subjects, eight had substitutions in the first loop, four had substitutions in the second loop and two had substitutions in both loops.

Table 2. The prevalence of MHR and RT mutations according to genotype/subgenotype

Genotype/ subgenotype	No.	MHR mutations	Prevalence, % (95% CI)	Escape mutations	Prevalence, % (95% CI)	Overlapping RT mutations	Prevalence, % (95% CI)
B	86	13	15.1 (7.5–22.7)	4	4.7 (0.2–9.2)	4	4.7 (0.2 to 9.2)
C	183	30	16.4 (11.0–21.8)	15	8.2 (4.2–12.2)	23	12.6 (7.8 to 17.4)
B2	70	11	15.7 (7.2–24.2)	4	5.7 (0.3–11.1)	3	4.3 (–0.5 to 9.1)
C2	108	14	13.0 (6.7–19.3)	5	4.6 (0.7–8.6)	11	10.2 (4.5 to 15.9)
C5	45	14	31.1 (17.6–44.6)	9	20.0 (8.3–31.7)	10	22.2 (10.1 to 34.3)

MHR, Major hydrophilic region; RT, reverse transcriptase; CI, confidence interval.

The difference in the prevalence of MHR mutations between C5 and C2 is $\chi^2 = 6.997$, $P < 0.05$. The differences in the prevalence of MHR mutations between B2 and C2, and B2 and C5, are $\chi^2 = 0.266$ and $\chi^2 = 3.817$, respectively (both $P > 0.05$). The differences in the prevalence of escape mutations between C5 and B2, and C5 and C2, are significant ($\chi^2 = 5.575$ and 9.027 , respectively, both $P < 0.05$) but the difference between B2 and C2 is not significant ($\chi^2 = 0.0008$, $P < 0.05$). The differences in the prevalence of overlapping RT mutations between C5 and B2, and C5 and C2, are $\chi^2 = 8.789$ and 3.887 , respectively, both $P < 0.05$, but the difference between B2 and C2 is $\chi^2 = 2.04$, $P > 0.05$.

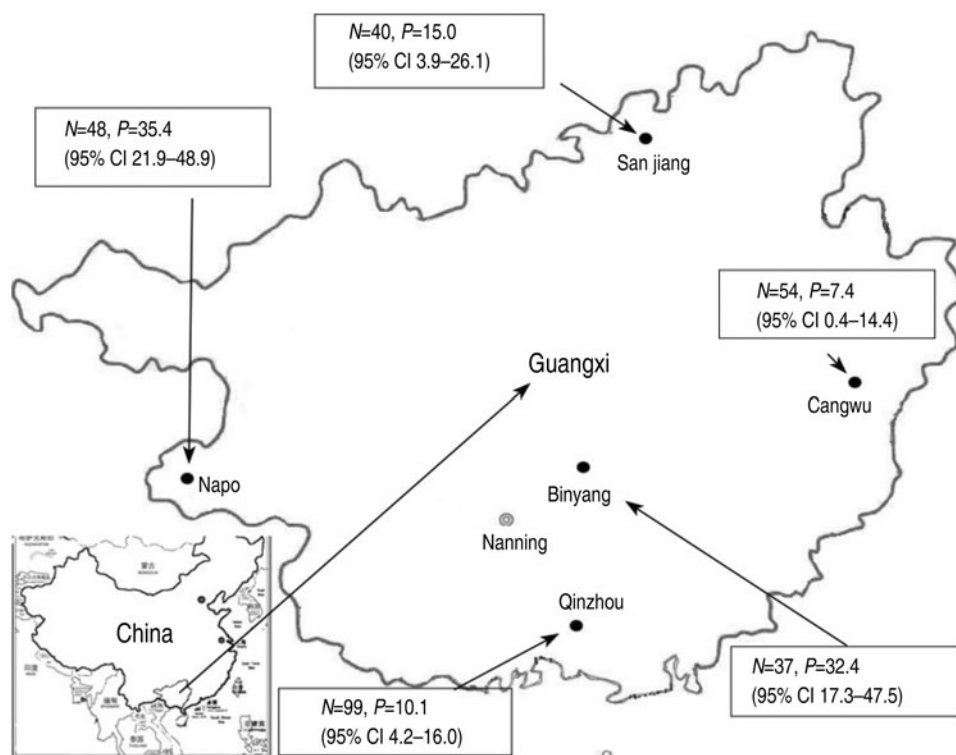


Fig. 1. The prevalence of major hydrophilic region (MHR) mutations in the various counties. *N*, Number of individuals; *P*, prevalence of MHR mutations; CI, confidence interval.

Substitutions Y100C [16], L109I [17] and M133 T [18] may lead to problems in diagnostic assays. The substitution I126S may lead to vaccine escape [18]. Substitutions T118 K [19], P120S and S143L [18] may lead to both problems in diagnostic assays and vaccine escape. Substitutions Q129H and G145A may lead to vaccine and HBIG therapy

escape [18]. Substitutions P120 T, T126N and G145R were involved in problems in diagnostic assays, vaccine escape and HBIG therapy escape [18]. Although substitution S132 F was suggested as an escape mutation by the Genafor/Arevir-geno2pheno drug resistance tool, its clinical effects remains unclear (Table 3).

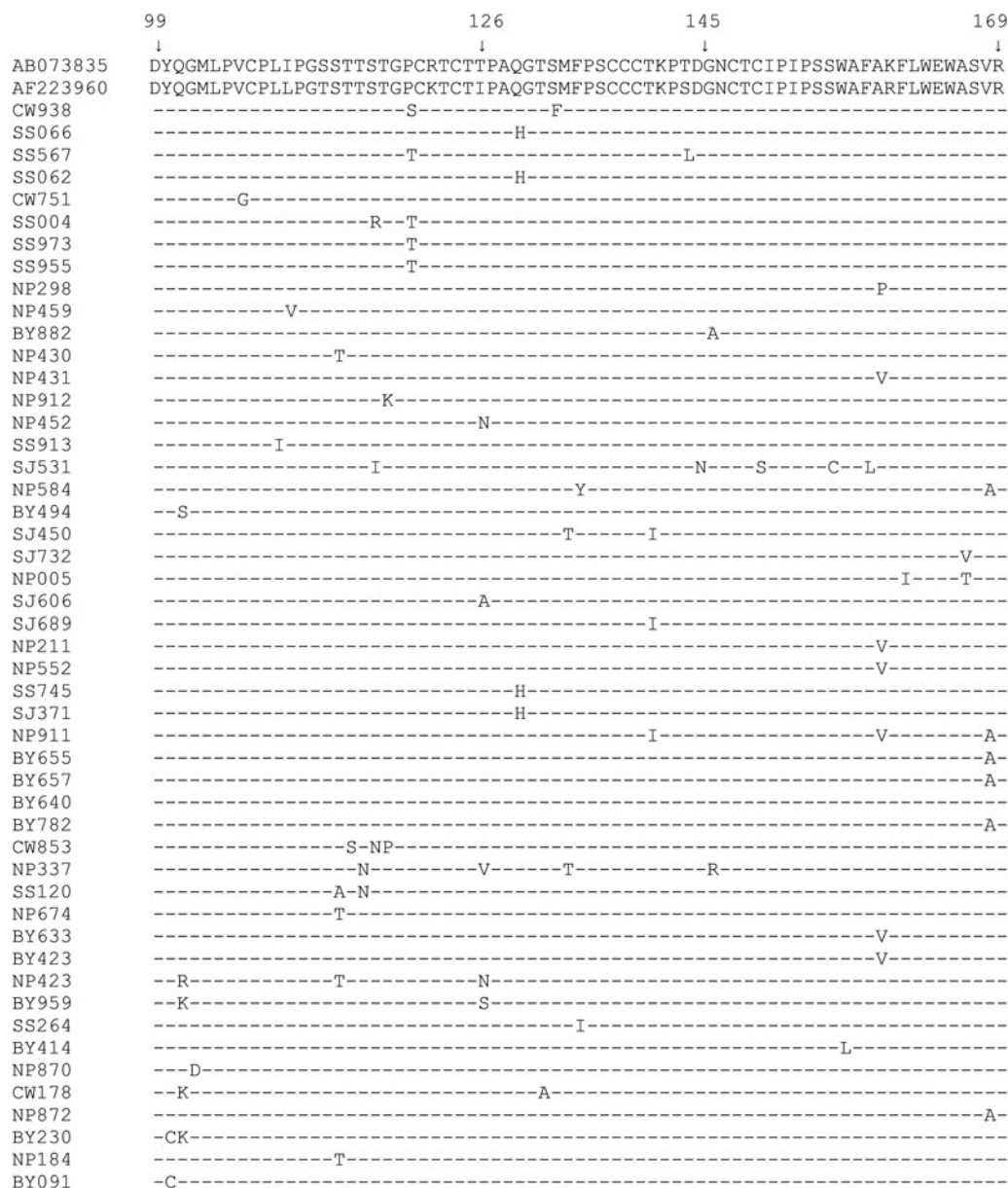


Fig. 2. Frequency and distribution of amino-acid substitutions in the major hydrophilic region of HBsAg.

Impact of mutations in the MHR on the overlapping polymerase region

Of the 43 mutations leading to amino-acid substitutions in the MHR, 72.1% (31/43) produced amino-acid changes in the overlapping polymerase (Table 3). The most frequent was rtI122N and T128N, corresponding to the substitutions S114 T and P120 T in the surface gene. Most of these changes (26/31) clustered in the A–B interdomain [2]. The remaining five changes were located in the B domain and one of them (rtI169N) is associated with resistance to entecavir when it is present with

rtL180M+rtM204 V/I. The overall prevalence of overlapping RT mutations was 10.1% (28/278, 95% CI 6.6–13.6). The differences in the prevalence of overlapping mutations between subgenotypes C5 and B2, and C5 and C2 are all significant ($\chi^2 = 8.789$ and 3.887 , respectively, all $P < 0.05$) but the difference between subgenotypes B2 and C2 is not significant ($\chi^2 = 2.04$, $P > 0.05$). The prevalence of overlapping RT mutations of subgenotype C5 is significantly higher, compared to that of subgenotypes B2 and C2 (Table 3). The data suggest that the prevalence of overlapping RT mutations corresponds with the prevalence of MHR mutations.

Table 3. *Escape mutations in the major hydrophilic region (aa 99–169) and amino-acid substitutions in the overlapping reverse transcriptase (aa 108–176)*

Major hydrophilic region		Reverse transcriptase	
Position (aa)	Escape mutations	Position (aa)	Mutations
100	Y100C (2*) (detection) [16†]	109	P109Q (3), P109S, S109 F
109	L109I (unclear) [17]	117	S117Y
118	T118 K (vaccine, detection) [19]	118	T118S
120	P120 T (4) [vaccine, therapy (IgG), detection], P120S (vaccine, detection) [18]	121	N121I
126	T126N (2) [vaccine, therapy (IgG), detection], I126S (vaccine) [18]	122	I122N (4), I122S
129	Q129H (4) [vaccine, therapy (IgG)] [18]	123	N123I
132	S132 F (unclear)	124	Y124Q (2)
133	M133 T (2) (detection) [18]	125	Q125H, Q125 K
140	T140I (3) [therapy (IgG)] [18]	126	H126P, H126Q
143	S143L (vaccine, detection) [18]	128	T128N (4), T128I
145	G145A [vaccine, therapy (IgG)], 145R [vaccine, therapy (IgG), detection] [18]	134	D134E (3), N134S, D134G
		139	N139S
		142	V142D
		151	F151Y
		152	G152E
		153	R153C, R153Q
		157	L157Q
		163	I163M
		164	L164 F
		167	R167P
		169	I169N
		174	G174D

aa, Amino acid; IgG, immunoglobulin.

* The number of individuals with the same mutation.

† Citation number.

Mutations in the S gene in subjects positive for both HBsAg and anti-HBs

Eighteen serum samples were positive for both HBsAg and anti-HBs. Only four of these had escape mutations. In order to search for mutations potentially associated with the co-existence of HBsAg and anti-HBs, the complete sequence of the S ORF was obtained from each of them. Various amino-acid substitutions were found, including S3N, T5A, T5I, V14A, A17 T, G18S, L21S, I25 F, W36R, F41S, L42I, T47A, T47P, S53L, P56S, T63I, P70R, C76G, L77R, L98 V, D99Y, S114 T, F170L, V177A, V194A, F200Y, S210N, C221Y, Y225S, N226H. However, according to current knowledge, none of these explains the co-existence of HBsAg and anti-HBs.

Multivariate logistic regression analysis of MHR status

Multivariable logistic regression analysis was performed to identify factors that affect the prevalence of MHR

mutations. The independent variables included gender, age, ethnicity, county, immunization history, alanine aminotransferase, subgenotypes, anti-HBs, HBeAg, anti-HBe and anti-HBc. Subgenotype and county are associated with MHR mutations. Subgenotype C5 was associated with MHR mutations [$P = 0.047$, odds ratio (OR) 2.4, 95% CI 1.0–5.9]. Individuals from Napo and Binyang counties have a higher prevalence of MHR mutations ($P = 0.002$, OR 4.8, 95% CI 1.8–12.8; $P = 0.007$, OR 4.8, 95% CI 1.6–14.7, respectively). Immunization did not influence the prevalence of MHR mutations (Table 4).

DISCUSSION

The major findings in this study are that the prevalence of MHR mutations varies with subgenotype. The prevalence of MHR mutations of viral surface antigen in subgenotype C5, including escape mutations and overlapping RT mutations, is significantly

Table 4. Logistic regression analysis of factors associated with MHR mutations

Variables	P value	HR	95% CI
Univariate analysis			
Sex			
Female*			
Male	0.533	1.238	0.633–2.420
Age, yr			
<15*			
≥15 to <50	0.584	0.760	0.285–2.027
≥50	0.896	1.076	0.357–3.240
Genotype			
Genotype B*			
Subgenotype C2	0.766	0.883	0.391–1.997
Subgenotype C5	0.055	2.355	0.981–5.655
Remainder of genotype C	0.248	0.401	0.085–1.892
County			
Binyang	0.022	3.320	1.189–9.267
Cangwu	0.391	0.553	0.143–2.141
Napo	0.001	4.770	1.885–12.069
Sanjiang	0.389	1.627	0.538–4.925
Qinzhou*			
Immunization			
No*			
Unknown	0.205	0.565	0.233–1.367
Yes	0.906	1.056	0.423–2.641
Ethnicity			
Han	0.684	0.814	0.301–2.197
Zhuang	0.606	1.312	0.467–3.683
Others*			
Alanine aminotransferase			
<40 IU*			
≥40 IU	0.452	1.305	0.652–2.611
Anti-HBs			
Negative*			
Positive	0.689	1.263	0.403–3.961
HBeAg			
Negative*			
Positive	0.243	1.498	0.760–2.953
Anti-HBe			
Negative*			
Positive	0.074	0.545	0.280–1.062
Anti-HBc			
Negative*			
Positive	0.417	2.354	0.298–18.605
Multivariate analysis			
Genotype			
Genotype B*			
Subgenotype C2	0.766	0.883	0.391–1.997
Subgenotype C5	0.047	2.433	1.011–5.857
Remainder of genotype C	0.248	0.401	0.085–1.892
County			
Qinzhou*			
Binyang	0.007	4.770	1.545–14.729
Cangwu	0.275	0.455	0.110–1.875
Napo	0.002	4.796	1.796–12.804
Sanjiang	0.330	1.878	0.528–6.681

HR, Hazard ratio; CI, confidence interval.

* The variable used for comparison.

higher than in other subgenotypes. Subgenotype C5 is common in Guangxi, China. More than two thirds of the MHR mutations led to amino-acid changes in the overlapping RT. The strength of this study is that the study design included different sampling locations, representing the whole of Guangxi province, and stratified random cluster sampling. The former allowed us to obtain information regarding MHR mutations for the whole province and compare the difference between counties, the latter allowed information regarding MHR mutations to be obtained for all age groups. A weakness of the study is that some individuals had been immunized against HBV, which may impact the MHR mutations, although the impact did not reach a significant level. Another weakness is that all subjects in the study are asymptomatic carriers and we have not included chronic HBV patients, so we do not know if the results of the study would be the same as in a clinical setting.

Understanding the prevalence of these mutations in a population is fundamental to immunoassay design and planning immunization programmes [20]. There have been some studies regarding the prevalence of naturally occurring MHR mutations [21–23]. However, most such studies were of clinical samples [20, 23–25] or blood donors [26, 27]. It has been reported that the prevalence of MHR mutations was associated with advanced liver disease [21, 28]. It also has been reported that age was a risk factor of MHR mutations [23, 24]. Therefore, data from clinical samples or blood donors may not reflect accurately the prevalence of MHR mutations in the general population. An elegant study conducted in China, which included the whole population, found that the prevalence of MHR mutations of genotype C is significantly higher than that of genotype B. Unfortunately, that study did not determine the prevalence rate of MHR mutations for subgenotypes [29]. In our study, we have the prevalence of MHR mutations for each subgenotype and found that the prevalence of MHR mutations differs significantly in subgenotypes.

It has been reported that there are some genotype-specific mutations in the HBV genome [30–32]. For example, the precore stop-codon mutation (A1896), accompanied by a C-to-T substitution at nt 1858 forming a base pair with it in the ϵ packaging signal, was found mainly in subgenotype C2 HBV. However, we could not find genotype-specific mutations in subgenotype C5. It is unclear why subgenotype C5 has such a high prevalence of MHR mutations.

In the present study, we also found that the prevalence of MHR mutations varies with geographical location. This could not be explained by genotypes or subgenotypes because the constituent of subgenotypes in each county is complicated and the sample size for each subgenotype in each county is too small to obtain statistical power. Therefore, the reason underlying the geographical difference in the prevalence of MHR mutations requires further study.

Amino-acid substitutions within the MHR may result in vaccine escape, failure of immunotherapy and lack of detection by commercial assays [33]. Nucleotide changes in the MHR may cause codon variations in the overlapping polymerase ORF [22] and alter susceptibility to antiviral therapy [8]. The mutations we found in this study may also lead to all these problems. Furthermore, we found that the prevalence of MHR mutations in subgenotype C5, including escape mutations and overlapping RT mutations, is significantly higher than in other subgenotypes. These findings are particularly useful for guiding the design of immunoassay, immunization programmes, immunotherapy and antiviral therapy, especially in regions where subgenotype C5 is common.

Our findings that the prevalence of MHR mutations varies with subgenotype may also be helpful in explaining some contradictory findings regarding genotypes and MHR mutations. For example, some studies reported that genotype D is a risk factor of MHR mutations [20] while others could not find this relationship [23]. When the subgenotypes were determined, it was found that the predominant subgenotype of the former is subgenotype D3, which has a higher prevalence of MHR mutations, while the predominant subgenotype of the latter is D7 [20, 23]. Therefore, these contradictory findings may be attributable to subgenotype.

It is believed that immunization may select mutations in the MHR [10]. However, immunization in the present study was not found to be a risk factor of MHR mutations. This may be attributed to the small sample size of the 1–15 years age group (the number of vaccinated children in this age group is only 20).

It has been reported that escape mutations in HBV are found predominantly in the first loop of the MHR, whereas those induced by immune pressure are located in the second loop [34]. In our study, about half of the escape mutations were located in the first loop. In addition, there were 14 subjects with

co-existing HBsAg and anti-HBs in the study. In two of them, no mutations were found in the S gene. All of the other subjects had mutations in the S gene but none are escape mutations according to the Genaford database. It has been reported that specific HBsAg MHR N-glycosylation mutations are implicated in HBV immune escape [35]. It is unclear whether any of these mutations have created sites for N-glycosylation.

Because the surface and polymerase ORFs overlap, nucleotide changes within the S ORF could introduce amino-acid substitution(s) into the polymerase [35]. However, the polymerase is more conserved and not all variants in MHR were accompanied by amino-acid substitutions in the polymerase [23]. This was confirmed by our results.

HBV C1 and C2 subgenotypes are thought to be the first or second most prevalent HBV genotype C strains in China [36–38]. However, our finding that subgenotype C5 is the second most common subgenotype of genotype C challenges the traditional view. As mentioned in the Results section, there remaining four samples could not be genotyped and 34 samples could not be subgenotyped based on the complete S gene. In the future, we plan to re-genotype the samples, based on the entire genomic sequences.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268815000242>.

ACKNOWLEDGEMENTS

We are indebted to staff members of Centers for Disease Prevention and Control of Cangwu, Qinzhou, Napo, Sanjiang and Binyang counties, Guangxi who assisted in recruiting the study subjects and to staff members of the Departments of Virology, Parasite and Chronic Non-communicable Diseases, Centre for Disease Prevention and Control of Guangxi, China for their help in recruiting the study subjects and handling the sera. This work was supported by the National Natural Science Foundation of China (grant no. 30960326/C150501).

DECLARATION OF INTEREST

None.

REFERENCES

1. **Tiollais P, et al.** The hepatitis B virus. *Nature* 1985; **317**: 489–495.
2. **Locarnini S, et al.** Molecular genetics of HBV infection. *Antiviral Therapy* 2010; **15** (Suppl. 3): 3–14.
3. **Desmond CP, et al.** A systematic review of T-cell epitopes in hepatitis B virus: identification, genotypic variation and relevance to antiviral therapeutics. *Antiviral Therapy* 2008; **13**: 161–175.
4. **Chen P, et al.** Computational evolutionary analysis of the overlapped surface (S) and polymerase (P) region in hepatitis B virus indicates the spacer domain in P is crucial for survival. *PLoS ONE* 2013; **8**: e60098.
5. **Coleman PF.** 2006. Detecting hepatitis B surface antigen mutants. *Emerging Infectious Diseases* 2006; **12**: 198–203.
6. **Chiou HL, et al.** Altered antigenicity of ‘a’ determinant variants of hepatitis B virus. *Journal of General Virology* 1997; **78**: 2639–2645.
7. **Chen HB, et al.** A novel hepatitis B virus mutant with A-to-G at nt551 in the surface antigen gene. *World Journal of Gastroenterology* 2003; **9**: 304–308.
8. **Ogura Y, et al.** 1999. Prevalence and significance of naturally occurring mutations in the surface and polymerase genes of hepatitis B virus. *Journal of Infectious Diseases* 1999; **180**: 1444–1451.
9. **Yamamoto K, et al.** Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. *Journal of Virology* 1994; **68**: 2671–2676.
10. **Purdy MA.** Hepatitis B virus S gene escape mutants. *Asian Journal of Transfusion Science* 2007; **1**: 62–70.
11. **Sticchi L, et al.** Epidemiology of HBV S-gene mutants in the Liguria Region, Italy: Implications for surveillance and detection of new escape variants. *Human Vaccines & Immunotherapeutics* 2013; **9**: 568–571.
12. **Fang ZL, et al.** Prevalence of hepatitis B virus infection in a highly endemic area of southern China after catch-up immunization. *Journal of Medical Virology* 2012; **84**: 878–884.
13. **Gunther S, et al.** 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *Journal of Virology* 1995; **69**: 5437–44.
14. **Hall TA.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999; **41**: 95–98.
15. **Tamura K, et al.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 2013; **30**: 2725–2729.
16. **Ireland JH, et al.** Reactivity of 13 in vitro expressed hepatitis B surface antigen variants in 7 commercial diagnostic assays. *Hepatology* 2000; **31**: 1176–82.
17. **Yong-Lin Y, et al.** Hepatitis B surface antigen variants in voluntary blood donors in Nanjing, China. *Virology Journal* 2012; **9**: 82.

18. **Avellón A, et al.** Frequency of hepatitis B virus 'a' determinant variants in unselected Spanish chronic carriers. *Journal of Medical Virology* 2006; **78**: 24–36.
19. **Mascagni P, et al.** Characterisation of an HBsAg mutant of hepatitis B virus (HBV) isolated from a dialysed patient involved in an occupational accident [in Italian]. *La Medicina del Lavoro* 2005; **96**: 231–237.
20. **Lazarevic I, et al.** Prevalence of hepatitis B virus MHR mutations and their correlation with genotypes and antiviral therapy in chronically infected patients in Serbia. *Journal of Medical Virology* 2010; **82**: 1160–1167.
21. **Song BC, et al.** Prevalence of naturally occurring surface antigen variants of hepatitis B virus in Korean patients infected chronically. *Journal of Medical Virology* 2005; **76**: 194–202.
22. **Sayiner AA, et al.** Naturally occurring MHR variants in Turkish patients infected with hepatitis B virus. *Journal of Medical Virology* 2008; **80**: 405–410.
23. **Kitab B, et al.** Hepatitis B genotypes/subgenotypes and MHR variants among Moroccan chronic carriers. *Journal of Infection* 2011; **63**: 66–75.
24. **Roque-Afonso AM, et al.** Viral and clinical factors associated with surface gene variants among hepatitis B virus carriers. *Antiviral Therapy* 2007; **12**: 1255–1263.
25. **Shi Y, et al.** Mutations in the major hydrophilic region (MHR) of hepatitis B virus genotype C in North China. *Journal of Medical Virology* 2012; **84**: 1901–1906.
26. **Liu Y, et al.** Molecular epidemiological study of hepatitis B virus in blood donors from five Chinese blood centers. *Archives of Virology* 2012; **157**: 1699–1707.
27. **Pourkarim MR, et al.** Evolutionary analysis of HBV 'S' antigen genetic diversity in Iranian blood donors: a nationwide study. *Journal of Medical Virology* 2014; **86**: 144–155.
28. **Panigrahi R, et al.** Subgenotype D5, BCP and MHR mutations in hepatic complications among hepatitis B virus infected patients from Orissa, India. *Infection, Genetics and Evolution* 2012; **12**: 1622–1629.
29. **Niu J, et al.** Variability of the S gene of hepatitis B virus in southeastern China. *Archives of Virology* 2010; **155**: 1951–1957.
30. **Yuan TT, et al.** The mechanism of natural occurrence of two closely linked HBV precore predominant mutations. *Virology* 1995; **211**: 144–156.
31. **Sung JJ, et al.** Genotype-specific genomic markers associated with primary hepatomas, based on complete genomic sequencing of hepatitis B virus. *Journal of Virology* 2008; **82**: 3604–3611.
32. **Mirandola S, et al.** Genotype-specific mutations in the polymerase gene of hepatitis B virus potentially associated with resistance to oral antiviral therapy. *Antiviral Research* 2012; **96**: 422–429.
33. **Echevarria JM, et al.** 2006. Hepatitis B virus genetic diversity. *Journal of Medical Virology* 2006; **78** (Suppl. 1): S36–42.
34. **Villet S, et al.** In vitro characterization of viral fitness of therapy-resistant hepatitis B variants. *Gastroenterology* 2009; **136**: 168–176.
35. **Yu DM, et al.** N-glycosylation mutations within hepatitis B virus surface major hydrophilic region contribute mostly to immune escape. *Journal of Hepatology* 2014; **60**: 515–522.
36. **Wang Z, et al.** Distribution and characteristics of hepatitis B virus genotype C subgenotypes in China. *Journal of Viral Hepatitis* 2007; **14**: 426–434.
37. **Wang Z, et al.** Hepatitis B virus genotypes and subgenotypes in China. *Hepatology Research* 2007; **37** (Suppl. 1): S36–41.
38. **Yuan J, et al.** Hepatitis B virus (HBV) genotypes/subgenotypes in China: mutations in core promoter and precore/core and their clinical implications. *Journal of Clinical Virology* 2007; **39**: 87–93.