To examine further the viral factors responsible for poor sensitivity to LAM, we analyzed the entire viral genome and identified two mutations associated with LAM sensitivity. The mutation in the enhancer I region (A1287G/C) was observed frequently in the good responder group as was the mutation in the precore region (G1896A). Enhancer I is responsible for upregulating transcription of the X mRNA and is also associated with viral replication. Although no transcription factor interacting with the region including nt 1287 has yet been identified, it is possible that the mutation impairs HBV replication efficiency and results in a good response to LAM. The precore mutation (G1896A) is known to be associated with HBeAg seroconversion. Recently, it was reported the precore mutation did not affect replication competence [20]. It is possible that host factors account for the good responsiveness to LAM of the patients infected with the precore mutant, but further studies are required to clarify the relationship between the presence of precore mutant and LAM sensitivity.

Regarding the deduced amino acid sequences, there was no significant difference between the two groups in the numbers of substitutions in the viral proteins, including preS1, preS2, surface, core, X, and polymerase. However, the number of substitutions in RT domain of the polymerase was found to be significantly higher in the good responders than in the poor responders. The polymerase is indispensable for HBV replication because the HBV genome is generated from the pregenomic RNA through reverse transcription, mediated by the RT domain of the polymerase. Because changes in the amino acid sequence of the RT domain might influence its enzymatic activity and reduce the efficiency of viral replication, it is possible that accumulation of substitutions in this domain might result in LAM-sensitive HBV. To clarify the correlation between the sensitivity to LAM and HBV polymerase sequence, further study is needed.

In conclusion, poor response to LAM in the first 24 weeks of therapy is one of the most significant parameters closely related to the occurrence of virological breakthrough. Full-genome sequence analysis of HBV suggested that point mutations in enhancer I and the precore region and the numbers of substitutions in the RT domain of the polymerase might be related to sensitivity to LAM treatment.

REFERENCES

1 Lai CL, Chien RN, Leung NW *et al.* A one-year trial of lamivudine for chronic hepatitis B. Asia Hepatitis Lamivudine Study Group. *N Engl J Med* 1998; 339(2): 61–68.

2 Dienstag JL, Schiff ER, Wright TL *et al.* Lamivudine as initial treatment for chronic hepatitis B in the United States. *N Engl J Med* 1999; 341(17): 1256–1263.

3 Allen MI, Deslauriers M, Andrews CW *et al.* Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Lamivudine Clinical Investigation Group. *Hepatology* 1998; 27(6): 1670–1677.

4 Perrillo R, Schiff E, Yoshida E *et al.* Adefovir dipivoxil for the treatment of lamivudine-resistant hepatitis B mutants. *Hepatology* 2000; 32(1): 129–134.

5 Lai CL, Rosmawati M, Lao J *et al.* Entecavir is superior to lamivudine in reducing hepatitis B virus DNA in patients with chronic hepatitis B infection. *Gastroenterology* 2002; 123(6): 1831–1838.

6 Tang S, Ho SK, Moniri K, Lai KN, Chan TM. Efficacy of famciclovir in the treatment of lamivudine resistance related to an atypical hepatitis B virus mutant. *Transplantation* 2002; 73(1): 148–151.

7 van Bommel F, Wunsche T, Mauss S *et al.* Comparison of adefovir and tenofovir in the treatment of lamivudine-resistant hepatitis B virus infection. *Hepatology* 2004; 40(6): 1421–1425.

8 Buti M, Sanchez F, Cotrina M *et al.* Quantitative hepatitis B virus DNA testing for the early prediction of the maintenance of response during lamivudine therapy in patients with chronic hepatitis B. *J Infect Dis* 2001; 183(8): 1277–1280.

9 Puchhammer-Stockl E, Mandl CW, Kletzmayr J *et al.* Monitoring the virus load can predict the emergence of drug-resistant hepatitis B virus strains in renal transplantation patients during lamivudine therapy. *J Infect Dis* 2000; 181(6): 2063–2066.

10 Ide T, Kumashiro R, Koga Y *et al.* A real-time quantitative polymerase chain reaction method for hepatitis B virus in patients with chronic hepatitis B treated with lamivudine. *Am J Gastroenterol* 2003; 98(9): 2048–2051.

11 Usuda S, Okamoto H, Iwanari H *et al.* Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. *J Virol Methods* 1999; 80(1): 97–112.

12 Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J Virol* 1995; 69(9): 5437–5444.

13 Fujiyama A, Miyanohara A, Nozaki C, Yoneyama T, Ohtomo N, Matsubara K. Cloning and structural analyses of hepatitis B virus DNAs, subtype adr. *Nucleic Acids Res* 1983; 11(13): 4601–4610.

14 Lok AS, Lai CL, Leung N *et al.* Long-term safety of lamivudine treatment in patients with chronic hepatitis B. *Gastroenterology* 2003; 125(6): 1714–1722.

15 Akuta N, Suzuki F, Kobayashi M *et al.* The influence of hepatitis B virus genotype on the development of lamivudine resistance during long-term treatment. *J Hepatol* 2003; 38(3): 315–321.

16 Huang YH, Wu JC, Chang TT *et al.* Analysis of clinical, biochemical and viral factors associated with early relapse after lamivudine treatment for hepatitis B e antigen-negative chronic hepatitis B patients in Taiwan. *J Viral Hepat* 2003; 10(4): 277–284.

17 Ciancio A, Smedile A, Rizzetto M, Lagget M, Gerin J, Korba B. Identification of HBV DNA sequences that are predictive of response to lamivudine therapy. *Hepatology* 2004; 39(1): 64–73.

18 Yuen MF, Sablon E, Hui CK, Yuan HJ, Decraemer H, Lai CL. Factors associated with hepatitis B virus DNA breakthrough

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

in patients receiving prolonged lamivudine therapy. *Hepatology* 2001; 34(4 Pt 1): 785–791.
19 Wolters LM, Hansen BE, Niesters HG *et al.* The influence of baseline characteristics on viral dynamic parameters in chronic hepatitis B patients treated with lamivudine. *J Hepatol* 2002; 37(2): 253–258.
20 Tacke F, Gehrke C, Luedde T, Heim A, Manns MP, Trautwein C. Basal core promoter and precore mutations in the hepatitis B virus genome enhance replication efficacy of Lamivudine-resistant mutants. *J Virol* 2004; 78(16): 8524–8535.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Journal of Medical Virology 79:683–693 (2007)

Analysis of the Complete Hepatitis B Virus Genome in Patients With Genotype C Chronic Hepatitis in Relation to HBeAg and Anti-HBe

KaiYu Zhang, Fumio Imazeki, Kenichi Fukai, Makoto Arai, Tatsuo Kanda, Rintaro Mikata, and Osamu Yokosuka*

Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, Japan

To investigate the relationship between viral factors and the development of chronic hepatitis B, the entire hepatitis B virus (HBV) genome of chronic carriers at different disease stages were analyzed. Eighty genotype C HBV carriers including 12 hepatitis B e antigen (HBeAg) positive asymptomatic carriers (Group A), 49 HBeAg positive patients with chronic liver diseases (Group B) and 19 anti-HBe positive patients with chronic liver diseases (Group C) were studied. HBV nucleic acid from serum samples was sequenced directly and compared with GenBank reference sequences HBV X01587 and M12906. On phylogenetic analysis, 76 cases were genotype C2. Of the 76 genotype C2 cases, the nucleotide and amino acid substitution rates in the precore/core region were significantly higher in Groups B and C than in Group A, also in Group C than in Group B. The nucleotide substitution rates in the full genome and the core promoter region were significantly higher in Group C than in Group A, also in group C than in Group B. The nucleotide and amino acid substitution rates in the X region were significantly higher in Group C than in Group A. The amino acid substitution rate in the pre-S2 region was significantly higher in Group C than in Group B. Deletion mutations were found mainly in Groups B and C. This whole genome analysis of HBV chronic carriers suggested that the nucleotide substitutions and deletions in HBV were closely associated with the pathogenesis of chronic HBV infection.
J. Med. Virol. 79:683–693, 2007.
© 2007 Wiley-Liss, Inc.

KEY WORDS: hepatitis B virus; full genome; chronic liver diseases

INTRODUCTION

Four hundred million people worldwide are infected chronically with hepatitis B virus (HBV) [Lok and McMahon, 2004]. The HBV genome comprises a partially double-stranded 3.2 kb DNA with four open-reading frames. Expression of the four transcripts is directed by the enhancer II/core promoter, large surface antigen promoter, major surface antigen promoter, and enhancer I/X promoter, respectively [Okamoto et al., 1987]. HBV reverse transcriptase lacks a proofreading function and is thus inherently error prone [Ganem and Varmus, 1987]. Hence, various substitutions may be observed throughout the HBV genome during long-term infection.

HBV may not be directly cytopathic and immune function is believed to play a key role in the development of chronic hepatitis B [Chisari and Ferrari, 1995]. The immune response to antigens encoded by HBV is responsible both for viral clearance and for disease pathogenesis during this infection. Chronic infection consists of three well-defined phases: an immune tolerance phase [hepatitis B e antigen (HBeAg) positive asymptomatic carrier], followed by an immune clearance phase with gradual disappearance of HBeAg, and a third non-replicative phase [Hadziyannis, 1995; Lee, 1997]. However, an undetermined proportion of anti-HBe positive carriers develop anti-HBe positive chronic liver diseases with a waxing and waning course of disease that can lead to cirrhosis [Brunetto et al., 1989; Hadziyannis and Vassilopoulos, 2001].

Substitutions in the precore/core region [Ehata et al., 1992] and the enhancer II/core promoter region [Honda et al., 1999] were often seen among patients with chronic liver diseases but not among HBeAg positive asymptomatic carriers. Although substitutions and deletions in the pre-S1/pre-S2/S region were reported in relation to the development of liver disease [Pollicino et al., 1995,

*Correspondence to: Osamu Yokosuka, Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-Ward, Chiba City, Chiba, Japan (260-8670). E-mail: yokosukao@faculty.chiba-u.jp

Accepted 16 February 2007
DOI 10.1002/jmv.20849
Published online in Wiley InterScience (www.interscience.wiley.com)

© 2007 WILEY-LISS, INC.

WILEY InterScience® DISCOVER SOMETHING GREAT

1997], the overall number of various substitutions in relation to the development of chronic hepatitis B is not well known.

Regarding studies on the full HBV genome in relation to chronic liver diseases, full length HBV genomes were determined in patients who were infected with different HBV genotypes [Huy et al., 2004], in serial serum samples obtained from an asymptomatic carrier who eventually developed hepatocellular carcinoma [Kajiya et al., 2002], in 14 patients with exacerbation of chronic hepatitis B [Liu et al., 2003] and in 40 patients with hepatocellular carcinoma [Takahashi et al., 1998]. However, the differences in the full length sequence of HBV among HBeAg positive asymptomatic carriers, patients with HBeAg positive chronic liver diseases and patients with anti-HBe positive chronic liver diseases are not well known. Therefore, in this study the full length sequence of HBV in these three phases were studied.

METHODS

Patients

Sera were obtained from 80 patients who were persistently positive for hepatitis B surface antigen (HBsAg) and who were followed at the Department of Medicine and Clinical Oncology, Chiba University Hospital, for at least 3 years. To examine the changes in the same genotype, only cases with genotype C were examined since more than 80% of the patients are infected with this genotype [Sumi et al., 2003]. They consisted of 12 HBeAg positive asymptomatic carriers and 68 patients with chronic liver diseases. The asymptomatic carriers showed alanine aminotransferase levels that were within the normal range on regular examinations performed every 3–6 months over 3–5 years. Among the 68 patients with fluctuating liver enzymes, 49 patients had HBeAg positive chronic liver diseases and 19 patients had anti-HBe positive chronic liver diseases. The HBV isolates from the asymptomatic carriers were placed in Group A, while the HBV isolates from the patients with HBeAg positive chronic liver diseases and anti-HBe positive chronic liver diseases were placed in Groups B and C, respectively. The HBV isolates in Group A were designated as asymptomatic carrier 1-asymptomatic carrier 12. The HBV isolates in Group B were designated as chronic liver diseases 1-chronic liver diseases 49, while those in Group C were designated as chronic liver diseases 50-chronic liver diseases 68. The clinical and laboratory data for these patients are summarized in Table I. None of the patients received any antiviral agents before collection of serum. None of these patients had a history of hepatitis C virus (HCV) or hepatitis D virus (HDV) coinfection. As to the cause of hepatitis, alcoholic liver disease and autoimmune hepatitis were ruled out in our patients. There was no clustering of similar or identical infections from a family group or dialysis center in these patients. All serum samples were collected and stored at −20°C until testing. Written informed consent was obtained from each patient.

TABLE I. Clinical and Laboratory Data of 12 HBV Asymptomatic Carriers and 68 Patients with HBV Associated Chronic Liver Diseases

	Gender	Age (yr)	WBC ($\times 10^3/\mu l$)	RBC ($\times 10^6/\mu l$)	PLT ($\times 10^3/\mu l$)	AST (IU/L)	ALT (IU/L)	ALB (g/dl)
Group A (HBeAg+ ASC)	M:F = 5:7	27.00 ± 9.12	5.78 ± 1.56	4.52 ± 0.54	204.50 ± 36.03	23.92 ± 5.87	14.33 ± 10.61	4.33 ± 0.34
Group B (HBeAg+ CLD)	M:F = 37:12	38.57 ± 9.00	5.51 ± 1.46	4.54 ± 0.47	170.69 ± 62.93	168.27 ± 330.57	217.24 ± 354.19	4.94 ± 5.71
Group C (anti-HBe+ CLD)	M:F = 13:6	45.68 ± 10.42	4.78 ± 1.27	4.54 ± 0.49	139.68 ± 57.62	77.00 ± 68.45	119.79 ± 144.30	4.25 ± 0.43
P (group A vs. group B)	*0.04*	*<0.001*	0.59	0.89	0.08	*<0.0001*	*<0.0001*	0.72
P (group A vs. group C)	0.26	*<0.001*	0.06	0.91	*<0.01*	*<0.0001*	*<0.0001*	0.59
P (group B vs. group C)	0.76	*<0.01*	0.06	1.00	0.07	0.24	0.25	0.60

HBV, hepatitis B virus; ASC, asymptomatic carrier; CLD, chronic liver diseases; HBeAg, hepatitis B e antigen; WBC, white blood cells; RBC, red blood cells; PLT, platelet; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALB, albumin.
The significance of differences (*P*-value) between the indicated groups was shown. $P < 0.05$ was indicated by italic.

HBV Markers

HBsAg, HBeAg and anti-HBe were detected by enzyme-linked immunosorbent assay (ELISA; Abbott Laboratory, North Chicago, IL). HBV genotypes were determined using patients' sera by ELISA (HBV Genotype EIA; Tokushu-Meneki Laboratory, Tokyo, Japan) [Usuda et al., 1999].

Amplification and Sequencing of the full HBV genome

A pair of synthetic oligonucleotide primers (p1 and p2) were prepared to amplify the full 3215 base pairs (bp) sequence of the HBV genome. In cases where the result of polymerase chain reaction (PCR) amplification of the 3215 bp HBV DNA was negative, second-round PCR with a pair of inner primers (p3 and p4) was performed to amplify a 3081 bp HBV DNA. PCR was performed according to the method described previously [Gunther et al., 1995] with minor modifications. In brief, viral DNA was extracted from 200 microliter (µl) of serum using a commercially available kit (QIAamp DNA Blood Mini Kit, QIAGEN, Hilden, Germany). PCR was performed with LA Taq (TaKaRa Bio Inc., Ohtsu, Japan) for 45 cycles in the first PCR (and 35 cycles in the second PCR). The conditions of PCR were an initial denaturation at 94°C for 5 min, followed by 94°C for 40 sec, 58°C for 40 sec, and 68°C for 3 min in a thermal cycler (TaKaRa Bio Inc., Ohtsu, Japan). The sequences between the joint points of the amplified 3215 bp and 3081 bp fragments, were amplified using the precore/core primers (p5 and p6 for the first PCR, p7 and p8 for the second PCR) [Sumi et al., 2003].

For sequencing, 50 µl of the PCR products was purified with a MinElute PCR purification kit (QIAGEN, Hilden, Germany) and 50–100 nanogram (ng) of the products was used for amplification with PCR primers in both directions. The nucleotide sequences were determined directly with an ABI PRISM 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence alignments were performed using the Auto Assembler of the GENETYX-MAC program, version 10.1 (GENETEX Corp., Tokyo, Japan). The positions and sequences of the PCR primers and sequencing primers are listed in Table II.

The HBV genotype C X01587 (subtype adr4) strain was selected as a reference strain, in which the reverse transcriptase domain (from nucleotide 130–1161) in the polymerase region is conserved [Stuyver et al., 2001]. All of the sequences were compared with that of X01587. The nucleotide (nt) and amino acid (aa) substitution rates were calculated (compared with the reference, the numbers of mutants in every isolate were counted in each region, and divided by the numbers of nucleotide and amino acid in the corresponding region). To confirm the results, all isolates were compared with another HBV genotype C reference strain M12906 (subtype adr), in which the gene organization of HBV DNA was found to be well conserved irrespective of subtype [Kobayashi and Koike, 1984], which had 2.7% nucleotide (87 nt) divergence in the full genome compared to X01587. A difference was reported to exist when the results

TABLE II. Positions and Sequences of the Primers Used for PCR Amplification and Sequencing

Position*	Nucleotide sequence (5′–3′)	Polarity
PCR primer		
p1:1821–1841	TTTTTCACCTCTGCCTAATCA	Sense
p2:1825–1806	AAAAAGTTGCATGGTGCTGG	Antisense
p3:1922–1940	GAATTTGGAGCTTCTGTGG	Sense
p4:1788–1771	GCCTACAGCCTCCTAGTA	Antisense
p5:1604–1623	TCGCATGGAGACCACCGTGA	Sense
p6:2076–2060	ATAGCTTGCCTGAGTGC	Antisense
p7:1653–1672	CATAAGAGGACTCTTGGACT	Sense
p8:1974–1957	GGAAAGAAGTCAGAAGGC	Antisense
Sequencing primer		
242–258	CAGAGTCTAGACTCGTGG	Sense
687–668	GGCACTAGTAAACTGAGCCA	Antisense
456–475	AAGGTATGTTGCCCGTTTGT	Sense
771–752	TACAGACTTGGCCCCCAATA	Antisense
668–687	TGGCTCAGTTTACTAGTGCC	Sense
1103–1086	GGCGAGAAAGTGAAAGCC	Antisense
1054–1073	ATGCCTTTATATGCATGTAT	Sense
1436–1418	GACGGGACGTAGACAAAGG	Antisense
1267–1285	CATACTGCGGAACTCCTAGC	Sense
2470–2453	TTATGAGTCCAAGGGATA	Antisense
2301–2320	CACCAAATGCCCCTATCTTA	Sense
2656–2639	GGATAGAACCTAGCAGGC	Antisense
2637–2656	ATGCCTGCTAGGTTCTATCC	Sense
3155–3138	CTTCCTGACTGCCGATTG	Antisense
3082–3099	CCTCAGGCTCAGGGCATA	Sense
475–456	ACAAACGGGCAACATACCTT	Antisense

*Nucleotide position based on the sequence of HBV genotype C (accession no. X01587). Primers p3 to p8 were also used for sequencing.

obtained using either of these two reference strains showed statistically significant differences from the sample sequences.

Phylogenetic Tree Analysis

A phylogenetic tree was constructed by the Unweighted Pair Group Method (UPGMA) to examine the heterogeneity of the viral sequence using the GENETYX-MAC program version 10.1 based on the entire genomic sequences of the 80 HBV isolates. The Genbank accession numbers of representative strains of subgenotypes C1–C4 of HBV were AF473543 in C1, X01587 and M12906 in C2 [Huy et al., 2004], (Norder et al. classified these three sequences as C2, C1 and C1, respectively [Norder et al., 2004]), X75656 in C3 and AB048704 in C4 [Norder et al., 2004; Tanaka et al., 2005].

Nucleotide sequences were multiple-aligned using GENETYX-MAC program version 10.1 and genetic distances were estimated using the Kimura two-parameter

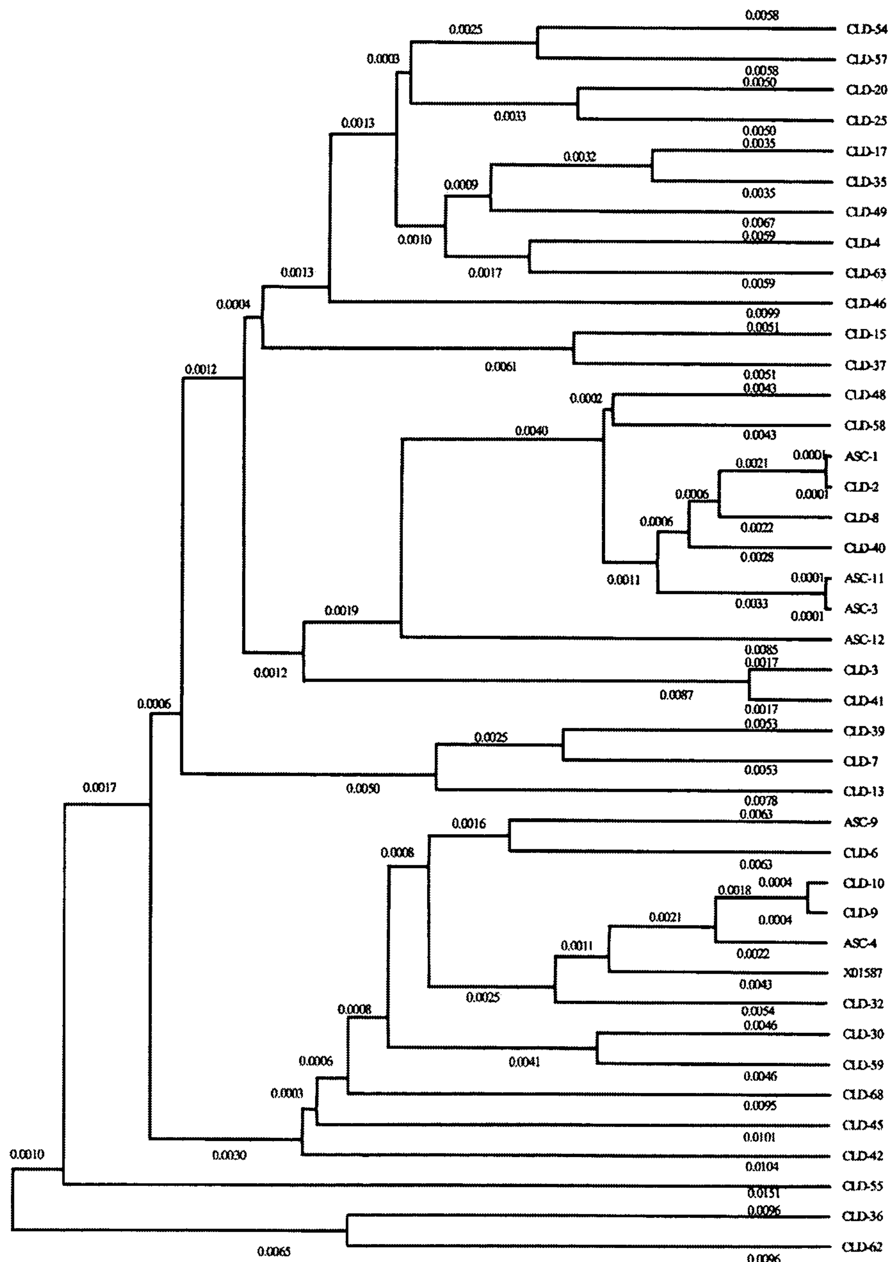

Fig. 1. Phylogenetic tree based on the entire genome sequences of 80 HBV isolates sequenced in this study. Genetic distances were estimated using the Kimura two-parameter matrix and the phylogenetic tree was constructed by UPGMA. Percentages at nodes represent levels of bootstrap support from 1000 resampled datasets. The HBV isolates in **Group A** were designated as ASC-1∼ASC-12, those in **Group B** as CLD-1∼CLD-49, and those in Group C as CLD-50∼CLD-68. Footprint was indicated at the bottom. The AF473543 (C1), X01587 and M12906 (C2), X75656 (C3) and AB048704 (C4) strains were representative strains of subgenotypes C1 to C4 of genotype C, and AY206389 was a genotype C HCC strain. ASC, asymptomatic carrier; CLD, chronic liver diseases.

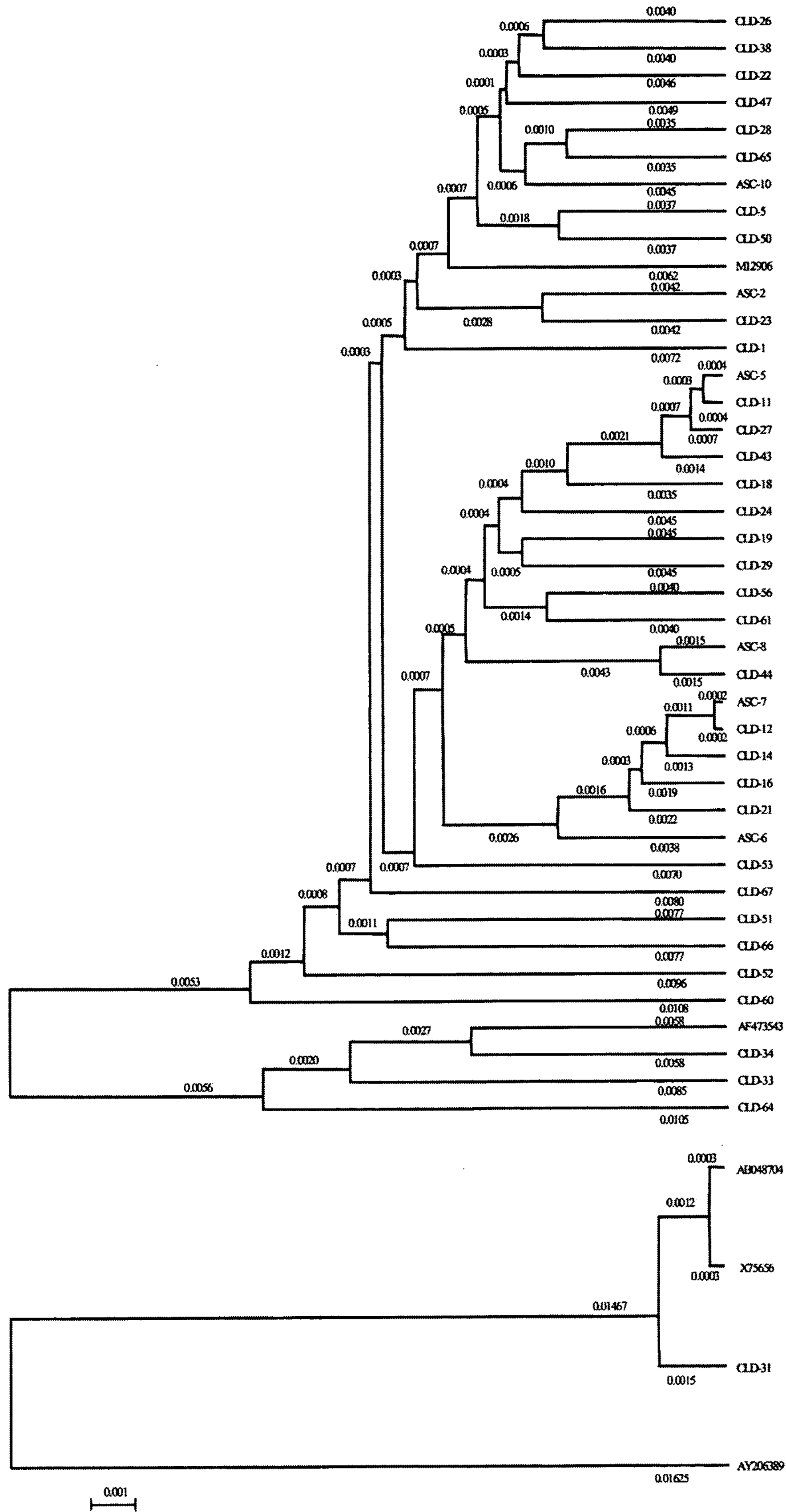

Fig. 1. (*Continued*)

matrix [Kimura, 1980]. The percentages at nodes represent levels of bootstrap support from 1000 resamplings of the data sets.

Statistical Analysis

All statistical analyses were performed with Statview 5.0 for Macintosh (SAS Institute Inc, Cary, NC). Statistical analyses were performed by Student's t -test and Fisher's exact probability test. Results are expressed as the mean $\pm$ standard deviation (SD). A two-tailed P values of less than 0.05 was considered to be significant.

RESULTS

Phylogenetic Tree Analysis of the 80 Full Length HBV Genomes

Phylogenetic tree analysis of the full sequence of the HBV genome of the 80 isolates resolved into three major groups: The first group was comprised of 76 isolates, while the second group was comprised of 3 isolates (chronic liver diseases 33, chronic liver diseases 34 and chronic liver diseases 64), and the third group was comprised of 1 isolate (chronic liver diseases 31) (Fig. 1). The 76 isolates in the first group were identified as genotype C2, the three isolates in the second group were identified as genotype C1 [Huy et al., 2004; Tanaka et al., 2005], and the isolate in the third group was homologous with AY206389 (genotype C, HCC strain).

Substitution Rates in Nucleotide and Amino Acid Sequences

The nucleotide and amino acid substitutions in the 76 genotype C2 isolates were analyzed. 12 Group A patients (12/12, 100%), 46 Group B patients (46/49, 94%) and 18 Group C patients (18/19, 95%) were infected with genotype C2 strain. The nucleotide substitution rates in the whole genome were compared among the three groups. The nucleotide substitution rate was significantly higher in Group C than in Group A ($P < 0.01$) (Tables III and IV). It was also significantly higher in Group C than in Group B ($P < 0.01$ when X01587 was used as the reference; $P < 0.05$ when M12906 was used as the reference) (Tables III and IV).

In the precore/core region, the nucleotide and the amino acid substitution rates significantly differed among Group A, Group B and Group C (Tables III–VI). As to substitutions in the core region, many non-synonymous substitutions in the N-terminal region of the core protein were detected, some of which were within the human lymphocyte antigen class I cytotoxic T lymphocytes epitope (codon 18 to 27), and additional substitutions were found in the mid-core region (data not shown).

In the pre-S1/ pre-S2/S regions, the nucleotide and amino acid substitution rates had no significant difference among the three groups compared with both reference HBV sequences except that the pre-S2 region amino acid substitution rate was significantly higher in

TABLE III. The Nucleotide Substitution Rates in the Full Genome and Various Regions of HBV in the Three Groups Compared with X01587

	Full genome (nt 0–3215)	Precore/core (nt 1814–2452)	Pre-S1 (nt 2848–3204)	Pre-S2 (nt 3205-154)	S (nt 155–835)	X (nt 1374–1838)	Pol (nt 2307-1620)	Enhancer I (nt 1060–1260)	Enhancer II (nt 1635–1714)	X promoter (nt 1230–1374)	Core promoter (nt 1519–1822)	S1 promoter (nt 2710–2800)	S2 promoter (nt 2960–3180)
Group A	20.97 ± 5.02*	12.52 ± 3.77*	24.98 ± 6.70*	25.41 ± 15.56*	7.10 ± 3.48*	19.00 ± 5.17*	22.14 ± 5.49*	40.22 ± 13.16*	26.04 ± 11.25*	45.92 ± 14.25*	17.60 ± 5.35*	21.06 ± 11.91*	35.82 ± 7.59*
Group B	21.35 ± 4.77*	16.13 ± 6.65*	26.25 ± 10.18*	24.52 ± 11.39*	7.57 ± 2.48*	19.31 ± 4.83*	21.64 ± 4.92*	38.61 ± 13.15*	20.92 ± 11.20*	43.48 ± 14.87*	21.74 ± 7.97*	18.63 ± 12.23*	37.87 ± 13.57*
Group C	25.73 ± 4.24*	24.69 ± 8.53*	30.66 ± 6.26*	27.78 ± 13.91*	9.06 ± 2.72*	23.18 ± 5.83*	24.43 ± 4.35*	41.74 ± 11.72*	25.69 ± 10.91*	46.11 ± 10.83*	27.78 ± 9.02*	25.03 ± 9.08*	41.98 ± 6.17*
(B vs. A)**	0.81	0.05	0.69	0.83	0.60	0.85	0.76	0.71	0.17	0.61	0.10	0.54	0.62
(C vs. A)**	*<0.01*	*<0.001*	*0.03*	0.67	0.10	0.05	0.21	0.74	0.93	0.97	*<0.01*	0.31	*0.02*
(C vs. B)**	*<0.01*	*<0.001*	0.09	0.34	*0.04*	*<0.01*	*0.04*	0.38	0.13	0.50	*0.01*	*<0.05*	0.22

S1 promoter, large surface antigen promoter; S2 promoter, major surface antigen promoter.
*The nucleotide (nt) substitution rates ($\times 10^{-3}$/nt; mean ± SD) in the full genome and the indicated regions of HBV were shown.
**The significance of differences (P-value) in the nucleotide substitution rate between the indicated groups was shown. $P < 0.05$ was indicated by italic.

TABLE IV. The Nucleotide Substitution Rates in the Full Genome and Various Regions of HBV in the Three Groups Compared with M12906

	Full genome (nt 0–3215)	Precore/core (nt 1814–2452)	Pre-S1 (nt 2848–3204)	Pre-S2 (nt 3205–154)	S (nt 155–835)	X (nt 1374–1838)	Pol (nt 2307–1620)	Enhancer I (nt 1060–1260)	Enhancer II (nt 1635–1714)	X promoter (nt 1230–1374)	Core promoter (nt 1519-1822)	S1 promoter (nt 2710–2800)	S2 promoter (nt 2960–3180)
Group A	17.94 ± 5.47*	8.34 ± 2.93*	27.78 ± 7.41*	17.28 ± 12.69*	5.87 ± 4.81*	25.27 ± 10.01*	19.05 ± 5.89*	37.72 ± 14.61*	25.00 ± 10.66*	43.65 ± 12.46*	21.19 ± 9.27*	16.48 ± 14.44*	39.97 ± 10.71*
Group B	20.84 ± 5.14*	12.04 ± 5.96*	29.59 ± 9.84*	16.17 ± 8.93*	6.96 ± 3.25*	30.06 ± 6.24*	21.38 ± 5.57*	37.85 ± 11.98*	31.52 ± 13.63*	50.58 ± 14.32*	28.02 ± 8.80*	16.72 ± 10.55*	43.87 ± 13.82*
Group C	23.78 ± 4.66*	20.08 ± 8.41*	31.59 ± 8.69*	22.02 ± 11.29*	8.24 ± 2.90*	33.09 ± 4.90*	22.65 ± 4.62*	37.04 ± 12.33*	36.81 ± 12.48*	44.22 ± 11.25*	33.52 ± 6.86*	20.76 ± 11.24*	42.99 ± 9.08*
(B vs. A)**	0.09	*0.04*	0.55	0.73	0.36	*0.04*	0.21	0.97	0.13	0.13	*0.02*	0.95	0.37
(C vs. A)**	*<0.01*	*<0.001*	0.22	0.29	0.10	*<0.01*	0.07	0.89	*0.01*	0.90	*<0.001*	0.37	0.41
(C vs. B)**	*0.04*	*<0.001*	0.46	*0.03*	0.15	0.07	0.39	0.81	0.16	0.10	*0.02*	0.18	0.80

S1 promoter, large surface antigen promoter; S2 promoter, major surface antigen promoter.
*The nucleotide (nt) substitution rates ($\times 10^{-3}$/nt; mean ± SD) in the full genome and the indicated regions of HBV were shown.
**The significance of differences (*P*-value) in the nucleotide substitution rate between the indicated groups was shown. $P < 0.05$ was indicated by italic.

the anti-HBe positive chronic liver diseases patients than in the HBeAg positive chronic liver diseases patients ($P = 0.02$) (Tables III–VI). When using X01587 as the reference strain, the pre-S2 region amino acid substitution rate between the HBeAg positive chronic liver diseases patients and anti-HBe positive chronic liver diseases patients showed a significant difference although the nucleotide substitution rate did not because non-synonymous amino acid substitutions were more frequently observed in the anti-HBe positive group than in the HBeAg positive group (53% vs. 44%). In the "α" determinant of the S region (nt 509–595, aa 119–147), the nucleotide and amino acid substitution rates showed no significant differences among the three groups (data not shown).

In the X region, the nucleotide and amino acid substitution rates were significantly higher in Group C than in Group A (Tables III–VI).

In the polymerase region, the nucleotide and amino acid substitution rates had no significant difference among Groups A, B and C (Tables III–VI). In the palm subdomain of the reverse transcriptase domain of HBV polymerase (nt 295–405, aa 402–438 and nt 642–924, aa 518–611) [Das et al., 2001], which encodes the catalytic sites, the nucleotide and amino acid substitution rates also showed no significant differences among Groups A, B and C compared with both reference HBV sequences (data not shown).

In the regulatory elements, the nucleotide substitution rate in the core promoter was significantly higher in Group C than in Group A ($P < 0.01$) and in Group C than in Group B ($P < 0.05$), but a significant difference was not seen between Group A and Group B compared with both reference HBV sequences (Tables III and IV).

The nucleotide substitution rates in the enhancer I, II and X promoter regions had no significant difference among the three groups (Tables III and IV). Similarly, in the large surface antigen promoter and in the major surface antigen promoter regions, the nucleotide substitution rates had no significant difference among the three groups compared with both reference HBV sequences (Tables III and IV).

The Nucleotide and Amino Acid Substitutions in Three Groups

Some sites of one point mutation with non-synonymous change were detected most frequently in anti-HBe chronic liver diseases patients compared the frequency of the mutation in each site among the other groups (Table VII).

In the core region, a change at codon 97 resulting in the amino acid substitution from isoleucine to leucine or phenylalanine was the most prominent and this was detected in none (0/12, 0%) of the Group A isolates, 7 (7/46, 15%) of the Group B isolates, and 8 (8/18, 44%) of the Group C isolates. In the pre-S1/pre-S2/S region, a noticeable substitution was found at codon 22 of the pre-S2 region, resulting in an amino acid substitution of phenylalanine to leucine. This substitution was found in

TABLE V. Amino Acid Substitution Rates in Various Regions of HBV in the Three Groups Compared with X01587

	Precore/core (213 aa)	PreS1 (119 aa)	PreS2 (55 aa)	S (226 aa)	X (155 aa)	Pol (843 aa)
Group A	2.75 ± 4.25*	24.51 ± 4.33*	33.33 ± 33.64*	15.49 ± 9.53*	29.57 ± 10.46*	28.17 ± 5.16*
Group B	10.87 ± 11.16*	32.52 ± 18.80*	28.46 ± 20.52*	15.10 ± 7.26*	35.48 ± 10.34*	28.57 ± 6.18*
Group C	26.21 ± 16.75*	32.68 ± 15.76*	42.42 ± 24.95*	20.40 ± 9.49*	43.73 ± 12.03*	29.46 ± 5.54*
(B vs. A)**	*0.02*	0.15	0.53	0.88	0.08	0.84
(C vs. A)**	*<0.001*	0.09	0.40	0.18	*<0.01*	0.53
(C vs. B)**	*<0.001*	0.97	*0.02*	*0.02*	*<0.01*	0.60

*The amino acid (aa) substitution rates ($\times 10^{-3}$/aa; mean ± SD) in the indicated regions of HBV were shown.
**The significance of differences (*P*-value) in the amino acid substitution rate between the indicated groups was shown. $P < 0.05$ was indicated by italic.

none (0/12, 0%) of the Group A, 3 (3/46, 6.5%) of the Group B, and 5 (5/18, 28%) of the Group C isolates. In the "α" determinant of the S region, a change at codon 126 was detected in 4 (4/12, 33%) of the Group A, 13 (13/46, 28%) of the Group B, and 5 (5/18, 28%) of the Group C isolates. Amino acid substitutions at codon 130 and/or 131 in the X region caused by A1762T and G1764A in the basic core promoter were detected in 3 (3/12, 25%) of the Group A, 35 (35/46, 76%) of the Group B, and 17 (17/18, 94%) of the Group C isolates. An amino acid substitution at codon 94 in the X region was detected in none (0/12, 0%) of the Group A, 8 (8/46, 17%) of the Group B, and 8 (8/18, 44%) of the Group C isolates. An amino acid substitution at codon 127 in the X region was detected in none (0/12, 0%) of the Group A, 11 (11/46, 24%) of the Group B, and 9 (9/18, 50%) of the Group C isolates.

Deletion Mutations in the Entire HBV Genome

In addition to nucleotide substitutions in the HBV genome, deletion mutations in the core promoter, core, pre-S1 and pre-S2 regions were detected. The deleted regions and the length of the deleted base pairs and corresponding protein regions of the 76 isolates are listed in Table VIII.

DISCUSSION

Phylogenetic tree analysis of 80 genotype C HBV sequences revealed the existence of three subgroups and the predominance of subgenotype C2 among the studied patients. This is concordant with the report that HBV genotype C strains could be classified into four subgroups, C1–C4 [Norder et al., 2004; Tanaka et al., 2005], and that genotype C2 was the main genotype in Far East Asia [Huy et al., 2004; Tanaka et al., 2005]. In order to compare HBV sequences within the same subgenotype, only cases of subgenotype C2 infection were analyzed in the current study.

Significant differences in the nucleotide substitution rates in the entire HBV genome were found, especially in the precore/core and core promoter regions between anti-HBe positive chronic liver diseases patients and HBeAg positive asymptomatic carrier or HBeAg positive chronic liver diseases patients when using both X01587 and M12906 as reference sequences for analysis. In this study, the patients with anti-HBe positive chronic liver diseases were older than those with HBeAg positive chronic liver diseases, and the patients with HBeAg positive chronic liver diseases were older than the HBeAg positive asymptomatic carriers. In Japan the transmission route of HBV is mainly perinatal infection; therefore, the longer the course of infection, the greater the number of mutations that accumulate. Significant differences in the amino acid substitution rates were also observed in the precore/core region among the three groups, in the X regions between the HBeAg positive asymptomatic carriers and anti-HBe positive chronic liver diseases patients, and in the pre-S2 region between the HBeAg positive chronic liver diseases patients and anti-HBe positive chronic liver diseases patients. These results might suggest that the substitutions are prone to develop in certain regions after HBeAg seroconversion.

It is well known that in patients with anti-HBe positive chronic liver diseases, the HBV genome contains a precore stop codon mutation that translationally suppresses the production of HBeAg [Carman et al., 1992]. Although the majority of patients with anti-HBe positive chronic liver diseases are believed to be infected

TABLE VI. Amino Acid Substitution Rates in Various Regions of HBV in the Three Groups Compared with M12906

	Precore/core (213 aa)	PreS1 (119 aa)	PreS2 (55 aa)	S (226 aa)	X (155 aa)	Pol (843 aa)
Group A	2.75 ± 4.25*	37.82 ± 8.40*	33.33 ± 33.65*	12.54 ± 17.17*	43.01 ± 17.03*	23.13 ± 5.73*
Group B	11.59 ± 11.19*	46.40 ± 19.06*	29.64 ± 17.71*	10.58 ± 8.57*	57.78 ± 13.04*	26.07 ± 6.88*
Group C	29.09 ± 19.30*	46.22 ± 17.35*	42.42 ± 24.95*	13.77 ± 7.87*	63.80 ± 11.69*	27.68 ± 5.55*
(B vs. A)**	*0.01*	0.14	0.60	0.58	*<0.01*	0.18
(C vs. A)**	*<0.001*	0.13	0.40	0.79	*<0.001*	*0.04*
(C vs. B)**	*<0.001*	0.97	*0.02*	0.18	0.09	0.38

*The amino acid (aa) substitution rates ($\times 10^{-3}$/aa; mean ± SD) in the indicated regions of HBV were shown.
**The significance of differences (*P*-value) in the amino acid substitution rate between the indicated groups was shown. $P < 0.05$ was indicated by italic.

TABLE VII. Nucleotide and Amino Acid Substitutions and Positive Rates in Several Sites of the Four Open-Reading Frames Compared with X01587 and M12906

	Group A	Group B	Group C	(B vs. A)*	(C vs. A)*	(C vs. B)*
1896 G-A (precore 28 W-stop)	0 (0%)	12 (26%)	10 (56%)	0.05	*0.001*	*0.04*
2189 A-C/T (core 97 I-L/F)	0 (0%)	6/1 (15%)	7/1 (44%)	0.33	*0.01*	*0.02*
53 T-C (pre-S$_2$ 22 F-L)	0 (0%)	3 (6.5%)	5 (28%)	1.00	0.07	*0.03*
531 T-C/G/A (S 126 I-T/S/N)	3/0/1 (33%)	9/4/0 (28%)	4/1/0 (28%)	0.73	1.00	1.00
1653 C-T (x 94 H-Y)	0 (0%)	8 (17%)	8 (44%)	0.19	*0.01*	0.05
1753 T-C/A (x 127 I-T/N)	0 (0%)	10/1 (24%)	6/3 (50%)	0.10	*<0.01*	0.07
1762 A-T (x 130 K-M)	3 (25%)	35 (76%)	17 (94%)	*<0.01*	*<0.001*	0.15
1764 G-A (x 131 V-I)	3 (25%)	35 (76%)	17 (94%)	*<0.01*	*<0.001*	0.15

*The significance of differences (P-value) in the nucleotide and amino acid substitution rate between the indicated groups was shown. $P < 0.05$ was indicated by italic.

with the precore variant [McMahon et al., 2001], this variant was also detected in approximately one-fourth of the HBeAg positive patients in the present study. This may suggest that either many patients tend to loose HBeAg or that patients who had reverted from the anti-HBe positive phase were included among the HBeAg positive chronic liver diseases patients.

Regarding nucleotide substitutions in the core region, clustering mutations in the mid-core region were detected, similar to the finding in a previous study [Ehata et al., 1992]. A nucleotide substitution at codon 97 leading to an amino acid substitution from isoleucine to leucine was most frequently observed. This change was previously revealed to enhance both the rate and extent of HBV capsid assembly compared with those of the wild-type protein [Ceres et al., 2004]. One of the reasons for the high rate of substitution at codon 97 may be due to the favorable replication capability acquired by this substitution [Suk et al., 2002].

A significant difference in the nucleotide substitution rate was found in the core promoter region between the asymptomatic carriers and chronic liver diseases patients due to frequent mutations at A1762T and G1764A in the chronic liver diseases patients. Although these two substitutions were reported to be involved in the production of HBeAg protein by modulating the transcription of the precore message and core message [Baumert et al., 1998], these substitutions were found in the majority of patients with chronic liver diseases irrespective of HBeAg status and, thus, it seems that these substitutions are related to the development of chronic liver diseases more than the status of HBeAg.

In the current study, deletions were found in the core promoter, core, pre-S1 and pre-S2 regions. The deletions in the core gene were located where B-cell epitopes [Colucci et al., 1988; Salfeld et al., 1989] and a helper T-cell epitope [Ferrari et al., 1991] exist. The deletion mutations in the pre-S2 region were located near the N-terminal portion where T- and B-cell epitopes exist [Tai et al., 2002]. Thus, these deletions seem to be caused by immune pressure similar to the nucleotide and amino acid substitutions [Akarca and Lok, 1995].

The deletions in the pre-S2 region overlapped the spacer region of the polymerase region which is not

TABLE VIII. Deletions in the Core Promoter, Core, pre-S1 and pre-S2 Regions of HBV Isolates

Sample number	Region*	Range**	Base pairs***	Related-protein Region****
CLD-6	Core promoter	1755–1774	20 bp	x 128–134
CLD-10	Core promoter	1755–1774	20 bp	x 128–134
CLD-12	Core promoter	1755–1774	20 bp	x 128–134
CLD-22	Core promoter	1756–1764	9 bp	x 128–130
CLD-25	Core promoter	1755–1774	20 bp	x 128–134
CLD-44	Core region	2137–2259	123 bp	c 108–149
CLD-49	Core region	2168–2290	123 bp	c 119–159
CLD-10	Pre-S1 region	2848–2865	18 bp	pre-S$_1$ 1–7, pol 181–187
CLD-12	Pre-S1 region	2848–2865	18 bp	pre-S$_1$ 1–7, pol 181–187
CLD-30	Pre-S1 region	2848–2865	18 bp	pre-S$_1$ 1–7, pol 181–187
CLD-21	Pre-S2 region	1–54	54 bp	pre-S$_2$ 4–22, pol 304–322
CLD-23	Pre-S2 region	22–54	33 bp	pre-S$_2$ 11–22, pol 311–322
CLD-29	Pre-S2 region	7–54	48 bp	pre-S$_2$ 6–22, pol 306–322
CLD-38	Pre-S2 region	49–57	9 bp	pre-S$_2$ 20–23, pol 320–323
CLD-54	Pre-S2 region	3205–3207	3 bp	pre-S$_2$ 1, pol 301
CLD-55	Pre-S2 region	30–53	24 bp	pre-S$_2$ 14–22, pol 314–322
CLD-56	Pre-S2 region	6–53	48 bp	pre-S$_2$ 6–22, pol 306–322
CLD-57	Pre-S2 region	40–54	15 bp	pre-S$_2$ 17–22, pol 317–322
CLD-61	Pre-S2 region	6–53	48 bp	pre-S$_2$ 6–22, pol 306–322

*Region in the HBV genome in which the deletion was located.
**Range of nucleotides that were deleted.
***Number of base pairs that were deleted.
****Protein region that were deleted, x, c, pre-S$_1$, pre-S$_2$, pol indicated X, core, pre-S1, pre-S2 and polymerase region, respectively.

essential for its enzymatic activity. Deletions in the pre-S2 region were seen more frequently in the group C anti-HBe positive phase (5/18, 28%) than in the group B HBeAg positive phase (4/46, 8.7%) in the current study. Interestingly, the amino acid substitution rate in the pre-S2 protein was significantly higher in the group C anti-HBe positive chronic liver diseases patients than in the group B HBeAg positive chronic liver diseases patients. It seems that amino acid substitutions in the pre-S2 protein occur more frequently after HBeAg seroconversion. The reason for this increase after HBeAg seroconversion is not known. It is suggested that generation of pre-S2 deletions and mutants after HBeAg seroconversion is a selection process that allows escape from immune pressure [Gerner et al., 1998].

Deletions in the core promoter and pre-S1 regions were found among the HBV isolates from chronic liver diseases patients. In the core promoter, the deletions encompassed either AT-rich region 1 (TA1) or 2 (TA2), or both, and a portion of TA3 which is located 20–35bp upstream of the precore mRNA start site. As these AT-rich sequences can bind to TATA-binding protein and play an important role in the function of core promoter [Chen et al., 1995], these deletions may affect the production of the viral proteins. A shift of the X gene frame and production of truncated X proteins were found in nearly all cases with a deletion in the core promoter. These truncated X proteins lack the C terminal domain that is required for the transactivation function of the X protein [Gunther et al., 1999]. These changes in its function as the core promoter or as a transactivator may reduce the expression of the viral proteins and may help them evade the immune system. The pre-S1 protein is also indispensable for viral replication, therefore, viruses with deletions in this region are unlikely to survive. An undetectable level of HBV that does not contain the pre-S1 deletion may exist and function as a helper virus. The role of deletion in these regions should be examined in future studies.

In conclusion, the nucleotide substitution rate and percentage of cases with deletions in the HBV genome were significantly higher in the chronic liver diseases patients than in the asymptomatic carriers. These substitutions and deletions may be caused by the interaction between the virus and host and seem to be closely associated with the development of chronic liver diseases.

REFERENCES

Akarca US, Lok AS. 1995. Naturally occurring hepatitis B virus core gene mutations. Hepatology 22:50–60.

Baumert TF, Marrone A, Vergalla J, Liang TJ. 1998. Naturally occurring mutations define a novel function of the hepatitis B virus core promoter in core protein expression. J Virol 72:6785–6795.

Brunetto MR, Oliveri F, Rocca G, Criscuolo D, Chiaberge E, Capalbo M, David E, Verme G, Bonino F. 1989. Natural course and response to interferon of chronic hepatitis B accompanied by antibody to hepatitis B e antigen. Hepatology 10:198–202.

Carman WF, Ferrao M, Lok AS, Ma OC, Lai CL, Thomas HC. 1992. Precore sequence variation in Chinese isolates of hepatitis B virus. J Infect Dis 165:127–133.

Ceres P, Stray SJ, Zlotnick A. 2004. Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L. J Virol 78: 9538–9543.

Chen IH, Huang CJ, Ting LP. 1995. Overlapping initiator and TATA box functions in the basal core promoter of hepatitis B virus. J Virol 69:3647–3657.

Chisari FV, Ferrari C. 1995. Hepatitis B virus immunopathogenesis. Annu Rev Immunol 13:29–60.

Colucci G, Beazer Y, Cantaluppi C, Tackney C. 1988. Identification of a major hepatitis B core antigen (HBcAg) determinant by using synthetic peptides and monoclonal antibodies. J Immunol 141:4376–4380.

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

Ehata T, Omata M, Yokosuka O, Hosoda K, Ohto M. 1992. Variations in codons 84-101in the core nucleotide sequence correlate with hepatocellular injury in chronic hepatitis B virus infection. J Clin Invest 89:332–338.

Ferrari C, Bertoletti A, Penna A, Cavalli A, Valli A, Missale G, Pilli M, Fowler P, Giuberti T, Chisari FV. 1991. Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. J Clin Invest 88:214–222.

Ganem D, Varmus HE. 1987. The molecular biology of the hepatitis B viruses. Annu Rev Biochem 56:651–693.

Gerner PR, Friedt M, Oettinger R, Lausch E, Wirth S. 1998. The Hepatitis B Virus seroconversion to anti-HBe is frequently associated with HBV genotype changes and selection of preS2-defective particles in chronically infected children. Virology 245:163–172.

Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. J Virol 69:5437–5444.

Gunther S, Fischer L, Pult I, Sterneck M, Will H. 1999. Naturally occurring variants of hepatitis B virus. Adv Virus Res 52:25–137.

Hadziyannis SJ. 1995. Viral hepatitis: an overview. Br J Clin Pract 49:267–268.

Hadziyannis SJ, Vassilopoulos D. 2001. Immunopathogenesis of hepatitis B e antigen negative chronic hepatitis B infection. Antiviral Res 52:91–98.

Honda A, Yokosuka O, Ehata T, Tagawa M, Imazeki F, Saisho H. 1999. Detection of mutations in the enhancer 2/core promoter region of hepatitis B virus in patients with chronic hepatitis B virus infection: comparison with mutations in precore and core regions in relation to clinical status. J Med Virol 57:337–344.

Huy TT, Ushijima H, Quang VX, Win KM, Luengrojanakul P, Kikuchi K, Sata T, Abe K. 2004. Genotype C of hepatitis B virus can be classified into at least two subgroups. J Gen Virol 85:283–292.

Kajiya Y, Hamasaki K, Nakata K, Nakagawa Y, Miyazoe S, Takeda Y, Ohkubo K, Ichikawa T, Nakao K, Kato Y, Eguchi K. 2002. Full-length sequence and functional analysis of hepatitis B virus genome in a virus carrier: a case report suggesting the impact of pre-S and core promoter mutations on the progression of the disease. J Viral Hepat 9:149–156.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Kobayashi M, Koike K. 1984. Complete nucleotide sequence of hepatitis B virus DNA of subtype adr and its conserved gene organization. Gene 30:227–232.

Lee WM. 1997. Hepatitis B virus infection. New Engl J Med 337:1733–1745.

Liu CJ, Chen PJ, Lai MY, Kao JH, Chang CF, Wu HL, Shau WY, Chen DS. 2003. A prospective study characterizing full-length hepatitis B virus genomes during acute exacerbation. Gastroenterology 124:80–90.

Lok AS, McMahon BJ. 2004. Chronic hepatitis B: update of recommendations. Hepatology 39:857–861.

McMahon BJ, Holck P, Bulkow L, Snowball M. 2001. Serologic and clinical outcomes of 1536 Alaska Natives chronically infected with hepatitis B virus. Ann Intern Med 135:759–768.

Norder H, Courouce AM, Coursaget P, Echevarria JM, Lee SD, Mushahwar IK, Robertson BH, Locarnini S, Magnius LO. 2004.

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

Okamoto H, Imai M, Kametani M, Nakamura T, Mayumi M. 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.

Pollicino T, Campo S, Raimondo G. 1995. PreS and core gene heterogeneity in hepatitis B virus (HBV) genomes isolated from patients with long-lasting HBV chronic infection. Virology 208: 672–677.

Pollicino T, Zanetti AR, Cacciola I, Petit MA, Smedile A, Campo S, Sagliocca L, Pasquali M, Tanzi E, Longo G, Raimondo G. 1997. Pre-S2 defective hepatitis B virus infection in patients with fulminant hepatitis. Hepatology 26:495–499.

Salfeld J, Pfaff E, Noah M, Schaller H. 1989. Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 63:798–808.

Stuyver LJ, Locarnini SA, Lok A, Richman DD, Carman WF, Dienstag JL, Schinazi RF. 2001. Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. Hepatology 33:751–757.

Suk FM, Lin MH, Newman M, Pan S, Chen SH, Liu JD, Shih C. 2002. Replication advantage and host factor-independent phenotypes attributable to a common naturally occurring capsid mutation (I97L) in human hepatitis B virus. J Virol 76:12069–12077.

Sumi H, Yokosuka O, Seki N, Arai M, Imazeki F, Kurihara T, Kanda T, Fukai K, Kato M, Saisho H. 2003. Influence of hepatitis B virus genotypes on the progression of chronic type B liver disease. Hepatology 37:19–26.

Tai PC, Suk FM, Gerlich WH, Neurath AR, Shih C. 2002. Hypermodification and immune escape of an internally deleted middle-envelope (M) protein of frequent and predominant hepatitis B virus variants. Virology 292:44–58.

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

Tanaka Y, Orito E, Yuen MF, Mukaide M, Sugauchi F, Ito K, Ozasa A, Sakamoto T, Kurbanov F, Lai CL, Mizokami M. 2005. Two subtypes (subgenotypes) of hepatitis B virus genotype C: a novel subtyping assay based on restriction fragment length polymorphism. Hepatol Res 33:216–224.

Usuda S, Okamoto H, Iwanari H, Baba K, Tsuda F, Miyakawa Y, Mayumi M. 1999. Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. J Virol Methods 80: 97–112.

Analysis of the complete hepatitis B virus genome in patients with genotype C chronic hepatitis and hepatocellular carcinoma

Kai Yu Zhang, Fumio Imazeki, Kenichi Fukai, Makoto Arai, Tatsuo Kanda, Rintaro Mikata and Osamu Yokosuka[1]

Department of Medicine and Clinical Oncology, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-Ward, Chiba City, Chiba 260-8670, Japan

(Received April 18, 2007/Revised July 27, 2007/Accepted August 5, 2007/Online publication September 20, 2007)

Hepatitis B virus (HBV) genotype C and the basic core promoter (BCP) mutations were reported to be associated with the development of hepatocellular carcinoma (HCC). In this study the full sequences of HBV genomes were analyzed in order to find the other predictors of HCC development. We determined the full sequences of HBV genomes in 24 genotype C carriers who developed HCC (HCC group) at the beginning of follow-up and at the time of HCC diagnosis, and 20 patients who did not develop HCC (non-HCC group) served as a control. The number of nucleotide and amino acid substitutions in most regions was higher in the HCC group than in the non-HCC group, and the following substitutions and deletions were found more frequently in the HCC group than in the non-HCC group: G1317A and T1341C/A/G in the X promoter region were detected in 13 and six of the HCC cases, four and none of the non-HCC cases, respectively; and pre-S2 deletion was detected in eight HCC and none of the non-HCC cases. Compared with the wild type X promoter, the mutant type X promoters, M1 (G1317A), M2 (T1341C), and M4 (T1341G) showed increases in activity of 2.3, 3.8, and 1.4 times, respectively, in HepG2 cells. Substitutions and deletion of nucleotides of the HBV genome, especially the pre-S2 deletion and G1317A and T1341C/A/G mutations may be useful markers for predicting the development of HCC. (*Cancer Sci* 2007; 98: 1921–1929)

Four hundred million people worldwide are chronically infected with HBV.[1] Infection with HBV leads to a wide spectrum of clinical presentations ranging from an ASC state to chronic hepatitis with progression to HCC in some patients. HCC is one of the major malignant diseases in the world, ranking as the fifth most common cancer, and it is believed that chronic HBV infection is a major global cause of HCC.[2]

HBV consists of four ORFs: the X, precore/core, pre-S/S, and Pol regions. Among these, the X and pre-S/S regions have been reported to function as transcriptional transactivators.[3–5] The domain between nt 221 and 573 in the pre-S/S region was termed the transactivator region, and its 3′-deleted form was shown to exert a transcriptional transactivator function.[6]

Although substitutions and deletions in the pre-S1/pre-S2/S region were reported in relation to the development of liver disease,[7,8] it remains unclear whether differences in the full length sequence of HBV exist at different stages of infection, particularly between patients who develop HCC and those who do not. Kajiya *et al.* analyzed the full sequence of serial serum samples obtained from a patient who underwent long-term follow-up from the time prior to development of symptoms to the chronic active hepatitis stage, and suggested that pre-S1 deletions and substitutions in the CP region may participate cooperatively in the progression of the disease.[9] Takahashi *et al.* studied the full length nucleotide sequence of HBV genome in sera from 40 Japanese patients with HBsAg positive HCC, and reported frequent deletions and missense mutations in the preS2 region.[10]

In the present study, the full length sequence of the HBV genome were analyzed in the sera of patients who did not have HCC at the beginning of follow-up and developed HCC thereafter, and in those of patients who did not develop HCC during follow-up served as the control group, in order to find the predicators of HCC development.

Materials and Methods

Patients. HBV sequences were examined in the sera of 24 patients who developed HCC (HCC group) at the beginning of follow-up (HCC 1B-24B) and at the time of HCC diagnosis (HCC 1A-24A). The control group (non-HCC group) consisted of 20 patients matched with respect to age, sex, and follow-up period who did not develop HCC, and were examined at the beginning of follow-up (non-HCC 1B-20B) and at the end of follow-up (non-HCC 1A-20A). The clinical and laboratory data for these patients are listed in Table 1.

All patients were followed at the Department of Medicine and Clinical Oncology, Chiba University. To compare differences in the same genotype, only cases with genotype C were examined since more than 80% of the patients were infected with this genotype.[11] All of the patients were negative for second-generation anti-HCV antibody (Ortho Diagnostics, Tokyo, Japan). All serum samples were collected and stored at –20°C until analysis. Written informed consent was obtained from each patient included in the study.

Amplification and sequencing of the full HBV genome. To amplify the full sequence of the HBV genome, two pairs of synthetic oligonucleotide primers (p1, p2 and p3, p4) were used. PCR was carried out according to the methods described previously,[12] with LA Taq (TaKaRa Bio Inc., Ohtsu, Japan) for 40 cycles in the first PCR and 35 cycles in the second PCR. The sequences between the joint points of the amplified PCR fragments were amplified using the precore/core primers (p5, p6 and p7, p8).[11]

For sequencing, 50 µL of the PCR product was purified with a MinElute PCR purification kit (QIAGEN, Hilden, Germany) and 50–100 ng of the products was used for sequencing with sequencing primers in both directions. The nucleotide sequences were determined directly with an ABI PRISM 3700 DNA

[1]To whom correspondence should be addressed.
E-mail: yokosuka@faulty.chiba-u.jp

Abbreviations: aa, amino acid; AP1, activator protein 1; ASC, asymptomatic carrier; ATF2, activating transcription factor 2; BCP, basic core promoter; C/EBP, CCAAT/enhancer-binding proteins; CLD, chronic liver diseases; CP, core promoter; CREB, cAMP response element-binding protein; DMEM, Dulbecco's modified minimal essential medium; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; LGE/mL: logarithm of genomic equivalent per milliliter; MEGA, molecular evolutionary genetics analysis; NRF1, nuclear respiratory factor 1; nt, nucleotide; ORFs, open reading frames; PCR, polymerase chain reaction; pHSA-R, polymerized human serum albumin receptor; Pol, polymerase; RLU, relative light units; RT, reverse transcriptase; SD, standard deviation; UPGMA, Unweighted Pair Group Method; X-PBP, X-promoter-binding protein.

doi: 10.1111/j.1349-7006.2007.00609.x
© 2007 Japanese Cancer Association

Table 1. Clinical and laboratory data of the 44 patients with and without HCC

	Gender	Age in year	HBeAg: anti-HBe	HBV-DNA (LGE/mL)	Follow-up time (months)	WBC($\times 10^3/\mu L$)	RBC($\times 10^6/\mu L$)	PLT($\times 10^3/\mu L$)	AST(IU/L)	ALT(IU/L)	ALB(g/dL)
Non-HCC group (B)	M:F = 14:6	41.35 ± 9.55	18:2	6.91 ± 2.90	–	5.09 ± 1.01	4.53 ± 0.45	182.55 ± 70.57	159.50 ± 415.58	214.55 ± 410.16	4.16 ± 0.34
Non-HCC group (A)	–	48.55 ± 8.56	18:2	6.31 ± 1.74	87.45 ± 36.43	5.27 ± 1.47	4.36 ± 0.48	183.15 ± 70.95	147.40 ± 173.11	209.55 ± 240.01	4.02 ± 0.70
HCC group (B)	M:F = 21:3	41.29 ± 10.51	14:10	6.21 ± 1.73	–	5.42 ± 1.62	4.54 ± 0.55	133.25 ± 63.93	89.96 ± 101.33	98.04 ± 92.79	4.17 ± 0.48
HCC group (A)	–	47.33 ± 9.14	15:9	5.48 ± 1.21	77.20 ± 41.86	4.68 ± 1.19	4.42 ± 0.55	118.33 ± 59.60	51.17 ± 36.86	49.50 ± 35.23	4.02 ± 0.57
P (non-HCC B *vs* HCC B)	0.26	0.99	*0.04*	0.13	–	0.43	0.94	*0.02*	0.43	0.18	0.90
P (non-HCC A *vs* HCC A)	–	0.65	*0.04*	0.39	0.40	0.15	0.74	*< 0.01*	*0.01*	*< 0.01*	1.00

The significance of differences (*P*-value) in the indicated groups is shown. $P < 0.05$ is indicated by italic type. Non-HCC group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC. ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HBeAg, hepatitis B e antigen; HCC, hepatocellular carcinoma; LGE/mL, logarithm of genomic equivalent per milliliter; PLT, platelet; RBC, red blood cells; WBC, white blood cells.

Table 2. Positions and sequences of the primers used for polymerase chain reaction (PCR) amplification and sequencing

Position[†]	Nucleotide sequence (5′–3′)	Polarity
PCR primer		
p1 : 1821–1841	TTTTTCACCTCTGCCTAATCA	Sense
p2 : 1825–1806	AAAAAGTTGCATGGTGCTGG	Antisense
p3 : 1922–1940	GAATTTGGAGCTTCTGTGG	Sense
p4 : 1788–1771	GCCTACAGCCTCCTAGTA	Antisense
p5 : 1604–1623	TCGCATGGAGACCACCGTGA	Sense
p6 : 2076–60	ATAGCTTGCCTGAGTGC	Antisense
p7 : 1653–1672	CATAAGAGGACTCTTGGACT	Sense
p8 : 1974–1957	GGAAAGAAGTCAGAAGGC	Antisense
Sequencing primer		
242–258	CAGAGTCTAGACTCGTGG	Sense
687–668	GGCACTAGTAAACTGAGCCA	Antisense
456–475	AAGGTATGTTGCCCGTTTGT	Sense
771–752	TACAGACTTGGCCCCCAATA	Antisense
668–687	TGGCTCAGTTTACTAGTGCC	Sense
1103–1086	GGCGAGAAAGTGAAAGCC	Antisense
1054–1073	ATGCCTTTATATGCATGTAT	Sense
1436–1418	GACGGGACGTAGACAAAGG	Antisense
1267–1285	CATACTGCGGAACTCCTAGC	Sense
2470–2453	TTATGAGTCCAAGGGATA	Antisense
2301–2320	CACCAAATGCCCCTATCTTA	Sense
2656–2639	GGATAGAACCTAGCAGGC	Antisense
2637–2656	ATGCCTGCTAGGTTCTATCC	Sense
3155–3138	CTTCCTGACTGCCGATTG	Antisense
3082–3099	CCTCAGGCTCAGGGCATA	Sense
475–456	ACAAACGGGCAACATACCTT	Antisense

[†]Nucleotide position based on the sequence of hepatitis B virus (HBV) genotype C (accession no. X01587). Primers p3 to p8 were also used for sequencing.

Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence alignments were carried out using the Auto Assembler of the GENETYX-MAC program, version 10.1 (GENETYX Corp., Tokyo, Japan). The positions and sequences of the PCR primers and sequencing primers are listed in Table 2.

We selected the HBV genotype C X01587 (subtype adr4) strain as a prototype, in which the reverse transcriptase (RT) region (from nt 130–1161) in the Pol ORF is conserved.[13] All of the sequences were compared with that of X01587. The number of nucleotide and amino acid substitutions was calculated. To confirm the results, we also compared all isolates with another HBV genotype C prototype, M12906 (subtype adr), in which the gene organization of HBV DNA was found to be well conserved irrespective of subtype,[14] and it had 2.7% nucleotide (87 nt) divergence in the full genome compared with X01587. A difference was reported to exist when the results obtained using either of these two reference strains showed statistically significant differences from the sample sequences.

The numbers of synonymous and non-synonymous substitutions in the ORF regions of HBV in the non-HCC and the HCC groups were estimated using MEGA software (download from www.megasoftware.net) compared with X01587 and M12906.

Phylogenetic tree analysis. To examine the heterogeneity of the viral sequence, a phylogenetic tree was constructed by the UPGMA using the GENETYX-MAC program version 10.1, based on the entire genomic sequences of the HBV isolates. The GenBank accession numbers of the representative genotype sequences C1-C4 of HBV were AF473543 in C1, X01587 and M12906 in C2,[15] (classified by Norder *et al.* as C2, C1 and C1, respectively[16]) X75656 in C3, and AB048704 in C4.[16,17]

doi: 10.1111/j.1349-7006.2007.00609.x
© 2007 Japanese Cancer Association

Table 3. Number of nucleotide substitutions in the full genome and various regions of hepatitis B virus (HBV) in the non-hepatocellular carcinoma (HCC) and HCC group

	Full genome (nt 0–3215)	pre-S1 (nt 2484–3204)	pre-S2 (nt 3205–154)	S (nt 155–835)	pre-C/core (nt 1814–2452)	X (nt 1376–1846)	Pol (nt 2307–1620)	Enhancer I (nt 1060–1260)	Enhancer II (nt 1635–1714)	X promoter (nt1230–1376)	core promoter (nt 1519–1822)	S1 promoter (nt 2710–2800)	S2 promoter (nt 2960–3180)
Non-HCC group (B)	68.35 ± 20.90	9.15 ± 3.73	4.40 ± 2.19	5.30 ± 2.16	10.30 ± 5.40	9.05 ± 3.25	54.55 ± 15.51	8.15 ± 3.05	1.70 ± 0.92	6.10 ± 1.99	5.30 ± 2.70	1.75 ± 1.12	7.95 ± 3.07
Non-HCC group (A)	71.85 ± 21.64	9.75 ± 3.58	4.65 ± 2.28	6.05 ± 2.19	12.55 ± 5.67	8.90 ± 2.56	56.30 ± 16.70	7.65 ± 2.70	1.60 ± 0.88	6.05 ± 1.88	5.35 ± 1.93	1.85 ± 1.50	8.20 ± 2.89
HCC group (B)	76.13 ± 11.59	10.26 ± 2.14	3.78 ± 1.93	5.74 ± 1.60	13.00 ± 5.75	9.69 ± 1.87	59.04 ± 8.23	8.09 ± 2.15	1.83 ± 0.98	7.26 ± 1.66	5.48 ± 1.44	1.78 ± 0.85	9.04 ± 1.36
HCC group (A)	80.26 ± 17.74	10.61 ± 2.87	4.39 ± 2.43	6.52 ± 3.87	14.52 ± 6.49	9.91 ± 2.71	61.65 ± 13.14	7.96 ± 1.99	1.70 ± 0.82	7.04 ± 1.92	5.43 ± 1.56	2.26 ± 1.39	9.61 ± 2.02
P (non-HCC B *vs* HCC B)	0.14	0.27	0.55	0.47	0.18	0.21	0.23	0.81	0.62	0.07	0.66	0.87	0.16
P (non-HCC A *vs* HCC A)	0.20	0.57	0.43	0.61	0.25	0.39	0.31	0.75	0.95	0.06	0.88	0.35	0.12

The significance of differences (*P*-value) in the indicated groups is shown. The number of nucleotide (nt) substitutions (mean ± SD) in the full genome and the indicated regions of HBV are shown. Non-HCC group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC.

Nucleotide sequences were multiple-aligned using GENETYX-MAC program version 10.1 and genetic distances were estimated using the Kimura 2-parameter matrix.[18] The percentages at nodes represent the levels of bootstrap support from 1000 resamplings of the data sets.

Preparation of plasmids with wild type and mutant type X promoters for transfection. pBluescript II SK+ plasmid containing full-length HBV DNA was used to make four variants by substituting the G of nt 1317 of the X promoter with A (M1), and the T of nt 1341 with C (M2), A (M3), or G (M4) using the Quickchange site-directed mutagenesis kit (Strategene). Mutations were verified by sequencing both strands of DNA in the regions of interest. Using the wild type and four mutant types of plasmid as the template, the region including enhancer I and X promoter was amplified by PCR using primers designed to introduce SacI and XhoI linker sequences (underlined) as follows: SacI CCGGAGCTCCCTGCGTTGATGCCTTTGTA and XhoI CCG-CTCGAGGAAACGATGTATATTTGCGG. To investigate differences in the activation of wild type and mutant type X promoters, experiments were carried out using an *in vivo* reporter system. In brief, the PCR products were cut by SacI and XhoI, and cloned into the SacI and XhoI sites of vector pGL4.10 (Promega Corp, Madison, WI, USA), which contains the firefly luciferase reporter gene without promoter, and hence, expression of the firefly luciferase gene in the reporter plasmid was controlled by the inserted enhancer I and X promoter. Transfection efficiency was monitored through the cotransfection of pRL-TK (Promega Corp, Madison, WI, USA), a control plasmid expressing *Renilla reniformis* (seapansy) luciferase driven by the herpes simplex virus thymidine kinase.

Cell culture and transfection and luciferase assays. Huh7 and HepG2 cells were purchased from Health Science Research Resources Bank (Osaka, Japan). Cells were maintained in DMEM (Sigma-Aldrich) supplemented with 100 U/mL penicillin, 100 µg/mL streptomycin, and 10% fetal calf serum at 37°C in a 5% CO_2 atmosphere.

Approximately 3×10^4 Huh7 and 2×10^4 HepG2 cells were plated in 24-well tissue culture plates (Iwaki Glass, Tokyo, Japan) 24 h prior to transfection. The transfection complexes containing a total of 0.2 µg of plasmids (0.12 µg of pGL4.10 and 0.08 µg pRL-TK) were added to each well using Effectene Transfection Reagent (Qiagen). Cells were harvested 24 h after transfection, and luciferase assays were carried out with the Dual–Glo™ Assay system (Promega). Firefly and seapansy luciferase activities were measured as relative light units (RLU) with a luminometer (AB-2200, ATTO. Tokyo, Japan).

Statistical analysis. All statistical analyses were carried out with Statview 5.0 software for Macintosh (SAS Institute Inc, Cary, NC). Statistical analyses were carried out by Student's *t*-test, Fisher's exact probability test, and Mann–Whitney *U*-test. Results were expressed as the mean ± standard deviation (SD). A two-tailed *P*-value of less than 0.05 was considered to be significant. Factors associated with the development of HCC were examined by multivariate Cox proportional regression analysis.

Results

Phylogenetic tree analysis of the full-length HBV genomes. Phylogenetic tree analysis of the full sequence of the HBV genome of the 44 isolates showed that all were genotype C2, except for one case (Case No. HCC 24B) that had nucleotide sequence identity with AY206389 (genotype C, HCC strain) (data not shown).

Substitutions in nucleotide and amino acid sequences. Nucleotide and amino acid substitutions were examined in the 43 genotype C2 isolates. The number of nucleotide substitutions in the full genome was 68.35 ± 20.90 and 76.13 ± 11.59 at the beginning of follow-up in the non-HCC group and HCC group,

© 2007 Japanese Cancer Association

Table 4. Number of amino acid substitutions in various regions of hepatitis B virus (HBV) in the non-hepatocellular carcinoma (HCC) and HCC groups

	Pre-C/core (213 aa)	Pre-S1 (119 aa)	Pre-S2 (55 aa)	S (226 aa)	X (155 aa)	Pol (843 aa)
Non-HCC group (B)	2.25 ± 2.81	3.70 ± 1.81	1.55 ± 1.28	3.45 ± 2.28	5.50 ± 2.24	24.35 ± 6.66
Non-HCC group (A)	3.90 ± 3.48	4.05 ± 1.85	1.60 ± 1.05	3.75 ± 1.33	5.65 ± 1.84	24.30 ± 5.93
HCC group (B)	3.70 ± 3.39	3.96 ± 1.66	1.87 ± 1.42	4.26 ± 1.76	5.91 ± 1.53	24.61 ± 3.68
HCC group (A)	4.09 ± 3.91	4.22 ± 1.93	2.17 ± 1.80	4.61 ± 2.52	6.13 ± 1.91	25.91 ± 5.20
P (non-HCC B *vs* HCC B)	0.17	0.87	0.43	0.19	0.47	0.97
P (non-HCC A *vs* HCC A)	0.90	0.84	0.25	0.21	0.33	0.35

The significance of differences (*P*-value) in the indicated groups is shown. The number of amino acid (aa) substitutions (mean ± SD) in the indicated regions of HBV is shown. Non-HCC group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC. Pol, polymerase.

Table 5. Mean numbers of synonymous and non-synonymous substitutions per nucleotide site in various regions of hepatitis B virus (HBV) in the non-hepatocellular carcinoma (HCC) and HCC groups compared with X01587

	Pre-C/core		Pre-S1		Pre-S2		S		X		Pol	
	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno
Non-HCC group (B)	0.0465	0.0046	0.0640	0.0145	0.0592	0.0144	0.0090	0.0070	0.0567	0.0158	0.0527	0.0138
Non-HCC group (A)	0.0510	0.0073	0.0692	0.0157	0.0636	0.0149	0.0084	0.0076	0.0538	0.0153	0.0542	0.0137
HCC group (B)	0.0570	0.0074	0.0927	0.0183	0.0096	0.0215	0.0052	0.0081	0.0579	0.0143	0.0617	0.0134
HCC group (A)	0.0615	0.0095	0.0841	0.0191	0.0118	0.0204	0.0070	0.0089	0.0563	0.0210	0.0630	0.0139
P (non-HCC B *vs* HCC B)	0.05	0.21	*0.001*	*<0.05*	*<0.001*	0.10	*0.01*	0.39	0.77	0.31	*0.03*	0.72
P (non-HCC A *vs* HCC A)	0.07	0.40	0.08	0.12	*< 0.001*	0.19	0.61	0.31	0.56	*<0.01*	0.07	0.87

The number of synonymous and non-synonymous substitutions per nucleotide site in the indicated regions of HBV is shown. The significance of differences (*P*-value) in the indicated groups is shown. $P < 0.05$ is indicated by italic type. Non-HCC group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC. Non-syno, non-synonymous; Pol, polymerase; syno, synonymous.

Table 6. Mean numbers of synonymous and non-synonymous substitutions per nucleotide site in various region of hepatitis B virus (HBV) in the non-hepatocellular carcinoma (HCC) and HCC groups compared with M12906

	Pre-C/core		Pre-S1		Pre-S2		S		X		Pol	
	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno	Syno	Non-syno
Non-HCC group (B)	0.0320	0.0046	0.0579	0.0230	0.0279	0.0152	0.0150	0.0052	0.0397	0.0251	0.0450	0.0152
Non-HCC group (A)	0.0355	0.0073	0.0628	0.0241	0.0326	0.0153	0.0144	0.0058	0.0401	0.0251	0.0459	0.0155
HCC group (B)	0.0379	0.0074	0.0713	0.0264	0.0094	0.0203	0.0112	0.0056	0.0361	0.0235	0.0457	0.0146
HCC group (A)	0.0427	0.0095	0.0693	0.0270	0.0119	0.0204	0.0130	0.0066	0.0374	0.0302	0.0501	0.0159
P (non-HCC B *vs* HCC B)	0.21	0.21	0.08	0.06	*<0.01*	0.17	*0.01*	0.78	0.29	0.36	0.85	0.48
P (non-HCC A *vs* HCC A)	0.18	0.40	0.40	0.22	*<0.01*	0.22	0.61	0.58	0.44	*<0.01*	0.34	0.70

The significance of differences (*P*-value) in the indicated groups is shown. $P < 0.05$ is indicated by italic type. The number of synonymous and non-synonymous substitutions per nucleotide site in the indicated regions of HBV is shown. Non-HCC group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC. Non-syno, non-synonymous; Pol, polymerase; syno, synonymous.

respectively ($P = 0.14$). The number was 71.85 ± 21.64 and 80.26 ± 17.74 at the end of follow-up in the non-HCC group and HCC group, respectively ($P = 0.20$) (Table 3). The average nucleotide substitution rates in the full genome were $0.90 \pm 0.96 \times 10^{-3}$/nt per year and $1.60 \pm 2.84 \times 10^{-3}$/nt per year in the non-HCC group and HCC group, respectively ($P = 0.26$). The number of amino acid substitutions in four ORFs was higher in the HCC group than in the non-HCC group, but there were no significant differences between them either at the beginning or at the end of follow-up (Table 4).

In the ORFs and regulatory elements, except for the pre-S2 and enhancer I regions, the HCC group tended to have more nucleotide substitutions, both at the beginning and at the end of follow-up, compared with the non-HCC group (Table 3).

We also estimated synonymous and non-synonymous substitutions in the ORF regions of HBV in the non-HCC and HCC groups. In almost every region, there were more synonymous and non-synonymous substitutions in the HCC group than in the non-HCC group at both the beginning and end of follow-up (Tables 5 and 6).

The pre-S2 region had more deletion mutants in the HCC group at both the beginning and end of follow-up (Fig. 1), and the deletion mutants were not included in the nucleotide or amino acid substitutions in the HCC group, so the actual number of nucleotide and amino acid substitutions (including the synonymous and non-synonymous substitutions) in the pre-S2 region were underestimated in this study.

The number of synonymous and non-synonymous substitution per nucleotide site was lower in the S region than in

doi: 10.1111/j.1349-7006.2007.00609.x
© 2007 Japanese Cancer Association

```
pre-S2 aa   1                                                     55
X01587      MQWNSTTFHQALLDPRVRGLYFPAGGSSSGTVNPVPTTASPISSISSRTGDPAPN
HCC-1B      X....................L...................T...F.........
HCC-2B      .............................................F.........
HCC-3B      T..T...S.....XXXXXXXXX.......................F.........
HCC-4B      .............................................F.........
HCC-5B      ...XXXXXXXXXXXXXXXXXXX.........A......V.....TF..I......
HCC-6B      .....................................S.......F.........
HCC-7B      .............................................F.........
HCC-8B      .............................................F.........
HCC-9B      .....XXXXXXXXXXXXXXXXX.......................F.........
HCC-10B     .............................................F.........
HCC-11B     ..........XXXXXXXXXXXX.......................F.........
HCC-12B     ................XXXXXX..........S............F.........
HCC-13B     .....................L.......................F.........
HCC-14B     .............................................F.........
HCC-15B     .............................................F.........
HCC-16B     .............................................F.........
HCC-17B     .............................................F.........
HCC-18B     .............................................F.........
HCC-19B     .....XXXXXXXXXXXXXXXXX.......................F.........
HCC-20B     .....................................S.......F.........
HCC-21B     V..............................A............TF..I..L...
HCC-22B     TH...XXXXXXXXXXXXXXXXX.......................F.........
HCC-23B     .............................................F.........

HCC-1A      X....................LR.............AA.......F..N......
HCC-2A      .............................................F.........
HCC-3A      ....C...QSXXXXXXXXXXXXH......................F.........
HCC-4A      .............................................F.........
HCC-5A      .....S.......................................F.........
HCC-6A      I....................L...............S.......F.........
HCC-7A      .............................................F.........
HCC-8A      .............................................F.........
HCC-9A      .....XXXXXXXXXXXXXXXXX.......................F.........
HCC-10A     .............................................F.........
HCC-11A     ..........XXXXXXXXXXXX.......................F.........
HCC-12A     .....................L.......................F.........
HCC-13A     .....................L.......................F.........
HCC-14A     .............................................F.........
HCC-15A     .............................................F.........
HCC-16A     .............................................F.........
HCC-17A     .............................................F.........
HCC-18A     .........KT..........L.......................F.........
HCC-19A     .....XXXXXXXXXXXXXXXXX.......................F.........
HCC-20A     .....................................S.......F.........
HCC-21A     V....................L.........A............TF..I..L...
HCC-22A     TH.XXXXXXXXXXXXXXXXXXX.......................F.........
HCC-23A     ...................XXX.......................F.........
```

```
pre-S2 aa   1                                                     55
X01587      MQWNSTTFHQALLDPRVRGLYFPAGGSSSGTVNPVPTTASPISSISSRTGDPAPN
non-HCC1B   .....................L.......................F.........
non-HCC2B   .............................................F.........
non-HCC3B   .........................................T...F.........
non-HCC4B   .............................................F.........
non-HCC5B   .............................................F.........
non-HCC6B   .............................................F.........
non-HCC7B   .............................................F.........
non-HCC8B   .............................................F.........
non-HCC9B   .............................................F.........
non-HCC10B  .............................................F.........
non-HCC11B  .............................................F.........
non-HCC12B  ...............................A............TF..I......
non-HCC13B  .............................................F.........
non-HCC14B  ...............................A............TF..I......
non-HCC15B  .............................................F.........
non-HCC16B  .....................L.........A.............F.......Q.
non-HCC17B  .............................................F.........
non-HCC18B  .....................L.................................
non-HCC19B  ........N......................A............TF..I......
non-HCC20B  .............................................F.........

non-HCC1A   .....................L.......................F.........
non-HCC2A   .............................................F.........
non-HCC3A   .....................L...................T...F.........
non-HCC4A   .............................................F.........
non-HCC5A   .............................................F.........
non-HCC6A   ...............K.............................F.........
non-HCC7A   ...T.........................................F.........
non-HCC8A   .............................................F.........
non-HCC9A   .............................................F.........
non-HCC10A  .............................................F.........
non-HCC11A  .............................................F.........
non-HCC12A  ...............................A............TF..I......
non-HCC13A  .............................................F.........
non-HCC14A  ...............................A............TF..I......
non-HCC15A  .............................E...............F.........
non-HCC16A  T....................L.........E.............F.......Q.
non-HCC17A  .............................................F.........
non-HCC18A  .....................L.......................F.........
non-HCC19A  .............................................F.........
non-HCC20A  .......................................................
```

Fig. 1. Deduced amino acid sequence encoded by the pre-S2 region in the 44 hepatitis B virus (HBV) isolates. Dots indicate amino acids that are identical to the respective amino acid in the pre-S2 region in X01587. X represents a deleted amino acid.

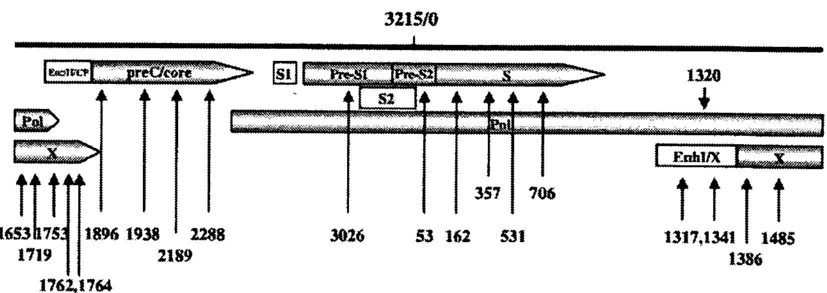

Fig. 2. Schematic diagram of the point mutations shown in Table 7. The number on the top of the full hepatitis B virus (HBV) genome refer to the *Eco*RI site (3215/0). Position of the open reading frames (ORFs) encoding precore/core (preC/core), pre-S1/pre-S2/S, polymerase (Pol), and X protein are shown by the gray rectangle and the arrow indicates the transcription direction. Enhancer I/X promoter (EnhI/X), enhancer II/core promoter (EnhII/CP), S1 promoter (S1) and S2 promoter (S2) are indicated by the open rectangle.

the other regions compared with both prototypes (Tables 5 and 6); therefore, the S region appeared to be strongly conserved.[19]

Nucleotide and amino acid substitutions in the non-HCC and HCC groups. The nucleotide and amino acid substitutions and positivity rates in amino acid sites of the four ORFs in the HCC and non-HCC groups are listed in Table 7 and Fig. 2.

At the end of follow-up, a comparison of the mutation frequency in the HCC group and the non-HCC group showed statistically significant substitution from adenine to guanine at 1317 (G1317A, hereafter) (14/23 *vs* 5/20, $P = 0.03$), T1341C/A/G (7/23 *vs* 0/20, $P = 0.01$), and C3026A/T (6/23 *vs* 0/20, $P = 0.02$) (Table 7). The frequency of these changes was also different at the beginning of follow-up between the HCC and the non-HCC groups (Table 7).

Furthermore, comparison of the frequency of mutations in the HCC group with that in the non-HCC group at the beginning of follow-up showed the following changes more frequent in the HCC group besides the above mentioned G1317A, T1341C/A/G (Table 7); G1764A with change of V to I at codon 131 in the X region (21/23 *vs* 11/20, $P = 0.01$), C2288A/G with changes of P to T/A at codon 130 in the core region (9/23 *vs* 2/20, $P = 0.04$), and T357C with a change of I to T at codon 68 in the S region (6/23 *vs* 0/20, $P = 0.02$) (Table 7).

© 2007 Japanese Cancer Association

Table 7. Nucleotide and amino acid substitutions and positive rates in several sites of the four open reading frames (ORFs) compared with X01587 and M12906

	Non-HCC group (B)	Non-HCC group (A)	HCC group (B)	HCC group (A)	*P* (non-HCC B *vs* HCC B)	*P* (non-HCC A *vs* HCC A)
1317 G-A	4 (20%)	5 (25%)	13 (56.5%)	14 (60.9%)	*0.03*	*0.03*
1341 T-C/A/G	0 (0%)	0 (0%)	3/2/1 (26.1%)	3/3/1 (30.4%)	*0.02*	*0.01*
1485 C-T (Xaa38 P-S)	2 (10%)	3 (15%)	4 (17.4%)	7 (30.4%)	0.67	0.29
1653 C-T (Xaa94H-Y)	5 (25%)	6 (30%)	6 (26.1%)	6 (26.1%)	1.00	1.00
1719T-G (Xaa116L-V)	6 (30%)	5 (25%)	0 (0%)	1 (4.3%)	*<0.01*	0.08
1753 T-C/A/G (Xaa127I-T/N/S)	5/1/1 (35%)	5/1/1 (35%)	6/4/0 (43.5%)	8/1/2 (47.8%)	0.76	0.54
1762 A-T (Xaa130K-M)	12 (60%)	12 (60%)	20 (87.0%)	20 (87.0%)	0.08	0.08
1764 G-A (X131V-I)	11 (55%)	15 (75%)	21 (91.3%)	21 (91.3%)	*0.01*	0.22
1896 G-A (preC28 W-stop)	4 (20%)	7 (35%)	9 (39.1%)	9 (39.1%)	0.20	1.00
1938 T-C (core13 V-A)	0 (0%)	0 (0%)	2 (8.7%)	4 (17.4%)	0.49	0.11
2189 A-C/T (core97 I-L/F)	3/0 (15%)	5/1 (30%)	5/0 (21.7%)	4/1 (21.7%)	0.70	0.73
2288 C-A/G (core130 P-T/A)	2/0 (10%)	4/1 (25%)	8/1 (39.1%)	6/2 (34.8%)	*0.04*	0.53
3026 C-A/T (pre-$S_1$60 A-V/E)	0 (0%)	0 (0%)	4/1 (21.7%)	5/1 (26.1%)	0.05	*0.02*
53 T-C (pre-$S_2$22 F-L)	2 (10%)	4 (20%)	2 (8.7%)	5 (21.7%)	1.00	1.00
Pre-S_2 deletion	0 (0%)	0 (0%)	8 (34.8%)	7 (30.4%)	*<0.01*	*0.01*
162 A-G (S3 N-S)	8 (40%)	8 (40%)	12 (52.2%)	12 (52.2%)	0.54	0.54
357 T-C (S68 I-T)	0 (0%)	1 (5%)	6 (26.1%)	5 (21.7%)	*0.02*	0.19
531 T-C/G (S126 I-T/S)	3/2 (25%)	3/3 (30%)	7/2 (39.1%)	7/2 (39.1%)	0.35	0.75
706 A-C (S184 V-A)	11 (55%)	11 (55%)	14 (60.9%)	13 (56.5%)	0.76	1.00
1386 G-C/A (X5 V-M/L)	4/1 (25%)	3/1 (20%)	4/4 (34.8%)	3/6 (39.1%)	0.33	0.20
1320 A-C (Pol743 K-N)	1 (5%)	1 (5%)	7 (30.4%)	7 (30.4%)	0.05	0.05

P-values for comparisons between the indicated groups are shown. $P < 0.05$ is indicated by italic type. Non-hepatocellular carcinoma (HCC) group (B): at the beginning of follow-up in patients who did not develop HCC. Non-HCC group (A): at the end of follow-up in patients who did not develop HCC. HCC group (B): at the beginning of follow-up in patients who developed HCC. HCC group (A): at the diagnosis of HCC in patients who developed HCC.

In the pre-S2 region, deletions between nt 3205–54 were detected in eight (34.8%) and seven (30.4%) of 23 cases in the HCC group at the beginning and the end of follow-up, respectively, but were not detected in the 20 non-HCC cases ($P = 0.0042$ and $P = 0.01$, respectively). The ranges of the deleted amino acids are listed in Fig. 1. All deletions ended at codon 22 in the pre-S2 region.

To further evaluate the effect of the above changes at the beginning of follow-up on the development of HCC, multivariate analysis (proportional hazards model) was carried out using the variables that were significantly associated with the presence of HCC at the beginning of follow-up as determined by Student's *t*-test and Fisher's exact probability test (Tables 1 and 7), showing only the G1317A mutation and HBeAg negativity were related with the development of HCC (Table 8).

Comparison of the activity of wild type and mutant type X promoters. The relative luciferase activity of firefly luciferase was compared among the wild type and mutant type X promoters. M1 (G1317A), M2 (T1341C), and M4 (T1341G) increased the activity of the X promoter 2.3, 3.8, and 1.4 times, respectively, compared with the wild type X promoter in HepG2 cells, but there were no significant changes in Huh7 cells (Fig. 3).

Table 8. Factors associated with the development of hepatocellular carcinoma (HCC): multivariate cox proportional analysis

Factor	Multivariate analysis		*P*-value
	Risk ratio	95% CI	
HBeAg			
Presence	1		
Absence	9.60	2.01–45.78	*0.005*
PLT(×10^3/μL)			
<100	1		
≥100	0.41	0.12–1.39	0.15
T357C change			
Absence	1		
Presence	1.85	0.25–13.97	0.55
Pre-S2 deletion			
Absence	1		
Presence	0.79	0.18–3.47	0.75
G1317A change			
Absence	1		
Presence	3.63	1.26–10.45	*0.02*
T1341C/A/G change			
Absence	1		
Presence	0.68	0.18–2.52	0.56
G1764A change			
Absence	1		
Presence	3.90	0.73–20.92	0.11
C2288A/G change			
Absence	1		
Presence	0.71	0.24–2.10	0.53

CI, confidence interval. $P < 0.05$ is indicated by italic type.

Discussion

This study showed that the number of nucleotide and deduced amino acid substitutions in the full genome and in each ORF region were higher in the HCC group than in the non-HCC group. The immune response to HBV-encoded antigens is responsible for both viral clearance and disease pathogenesis during HBV infection. The dominant cause of viral persistence during chronic HBV infection is the development of a weak antiviral immune response to the viral antigen.[20] The nucleotide and deduced amino acid substitutions in the HBV genome will change the target epitopes of some HBV proteins, may help

doi: 10.1111/j.1349-7006.2007.00609.x
© 2007 Japanese Cancer Association

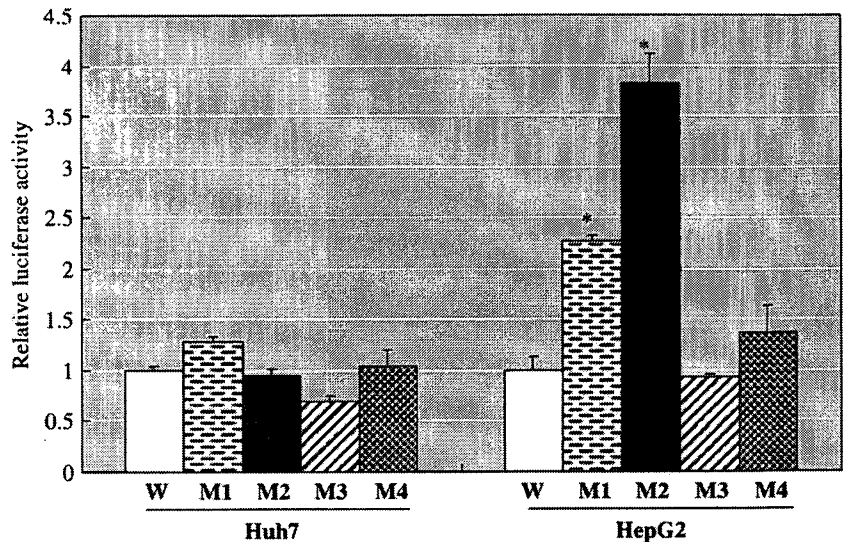

Fig. 3. Activation of the wild type and mutant type X promoters, M1 (G1317A), M2 (T1341C), M3 (T1341A), and M4 (T1341G) in Huh7 and HepG2 cells. [Correction added after online publication 23 October 2007: '(T1314C), M3 (T1314A), and M4 (T1314G)' changed to '(T1341C), M3 (T1341A), and M4 (T1341G)'.] The observed firefly luciferase activities (mean ± SD) were normalized to the mean *Renilla reniformis* luciferase activities. *Indicates mean values in mutant type X promoters that are significantly different from that of the wild type X promoter ($P < 0.05$).

HBV to evade host-immune surveillance, and thus may contribute to the progression to HCC. The X region in particular had significantly different non-synonymous substitutions between the non-HCC group and the HCC group at the end of follow-up (Tables 5 and 6), so these non-synonymous substitutions in HBx may abolish the antiproliferative and apoptotic effect of HBx, thereby causing uncontrolled growth and multiple-step hepatocarcinogenesis.[21]

Furthermore, we found several mutations and deletions in the HBV genome more frequently in the HCC group than in the non-HCC group at the beginning of follow-up. These included G1317A and T1341C/A/G mutation in the X promoter and pre-S2 deletion. Nt 1317 and 1341 were overlapped with the Pol region but these two mutations did not change the amino acid sequence of Pol. The results of the reporter assay showed that the mutant promoters with G1317A and T1341C/G retained higher X promoter activity in HepG2 cells. We therefore speculate that these point mutations might increase the synthesis of HBx mRNA and could be associated with the development of HCC, since HBx has been shown to function as a transcriptional transactivator of a variety of viral and cellular promoter and/or enhancer elements,[22] and the expression of HBx was also reported to be closely related to the pathogenesis of HCC.[23]

X-promoter-binding protein (X-PBP) specially interacts with a 16-bp sequence (nt 1228–1243 in our study) which has promoter activity in X gene transcription. This target sequence is located in the 58-bp sequence (nt 1211–1268 in our study), which could be enhanced by the HBV enhancer.[24] Recently, NRF1 was reported to specially bind the 21-bp (including the 16-bp sequence above) minimal promoter and to positively contribute to transcription of the X gene.[25] In addition to the X-PBP and NRF-1 binding sites, other binding sites interact with the transcription factors C/EBP, CREB/ATF2, AP1 located about 130 bp upstream of the X protein-encoding region.[26] In our study, nt 1317 and nt 1341 were not located in the aforementioned binding sites that interact with X-PBP and other transcription factors, but changes at nt 1317 and nt 1341 influenced the activity of the X promoter in HepG2 cells, indicating that changes at these two sites might influence X gene transcription to some degree. However, the mutant promoter with T1341A did not change X promoter activity, which was also detected in 8.7% (2/23) of HCC group patients at the beginning of follow-up, and X promoter activity did not differ between wild type and mutant type X promoters in Huh7 cells; effects of these mutant types of X promoters warrant further study.

Since more pre-S2 deletion isolates were detected in the pre-follow up sera of patients who developed HCC, it is likely that the deletion mutant in pre-S2 is predictive of HCC development. Naturally occurring pre-S2 deletion has been found in serum samples from HBV carriers.[27,28] Huy *et al.* studied 387 HBV DNA positive serum samples of individuals from 12 countries, and pre-S2 deletion was observed in 33 samples (8.5%), and was found exclusively at the 5′-terminus of pre-S2 region, and pre-S2 mutants were detected at an even higher rate in HCC cases (34.7%).[29] Takahashi *et al.* examined sera from 40 cases of HBV-related HCC and found that 10 of 40 (25%) isolates had pre-S2 deletions, and mutations of codon 22 were detected in 30% of cases.[10] These results are in accordance with the present study, which showed the detection of pre-S2 deletions in seven of 23 (30.4%) isolates in the HCC group. A noteworthy finding was that all deletions detected in the pre-S2 region ended at amino acid 22 of the pre-S2 region in the current study (Fig. 1).

In the pre-S2 region, pHSA-R (aa 3–16),[30] has been identified, against which neutralizing antibodies develop.[31] Previous studies have shown that HLA-restricted B- and T-cell epitopes of HBV exist in the middle envelope proteins, and middle envelope proteins with a pre-S2 internal deletion were not recognized *in vitro* by a putative neutralizing antiserum, suggesting that these variants can evade immune recognition *in vivo*,[32] and change the course of HBV infection.[33] The deleted region in this study mainly included the pHSA-R and B-cell epitope. Chen *et al.* reported that the detection of pre-S deletions overlapping with the pHSA-R region were detected in 57.1% of HCC patients with genotype C, and most of the deletions ended at nt 60.[34] The pre-S2-deleted mutant has been reported to display tumor-promoting phenotypes in Huh7 cells, for example, enhanced proliferation and clonal expansion abilities,[35] and to cause strong oxidative stress and overall genomic instability,[36] the induced genomic instability surely enhances HCC development,[36] and this may be related with the oncogenic properties of this mutant virus.[31] Furthermore, integration of the truncated pre-S/S gene has been proposed to enhance the development of HCC by expressing a transactivating capacity.[37]

It has been suggested that older age, liver cirrhosis and positive HBeAg are important risk factors for HCC.[38–40] Recently,

© 2007 Japanese Cancer Association