Epidemiological and Virological Characteristics of 2 Subgroups of Hepatitis B Virus Genotype C

Henry L. Y. Chan,1 Stephen K. W. Tsui,2 Chi-Hang Tse,1 Eddie Y. T. Ng,1 Thomas C. C. Au,2 Lilly Yuen,4 Angeline Bartholomeusz,4 Kwong-Sak Leung,2 Kin-Hong Lee,3 Stephen Locarnini,4 and Joseph J. Y. Sung1

Departments of 1Medicine and Therapeutics, 2Biochemistry, and 3Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, People’s Republic of China; 4Victorian Infectious Disease Research Laboratory, Melbourne, Australia

Background. We aimed to investigate the characteristics of hepatitis B virus (HBV) genotype C subgroups in Hong Kong and their relationship with HBV genotype C in other parts of Asia.

Methods. Full-genome nucleotide sequences of 49 HBV genotype C isolates from Chinese patients with chronic hepatitis B were compared with the sequences of 69 HBV genotype C isolates and 12 non–genotype C isolates in the GenBank database. Phylogenetic analysis was performed to define the subgroups of HBV genotype C on the basis of >4% heterogeneity of the entire HBV genome.

Results. HBV in 80% of patients in Hong Kong belonged to a subgroup predominantly found in Southeast Asia (Vietnam, Thailand, Myanmar, and southern China) designated as HBV genotype “Cs,” and HBV in the remaining 20% of patients belonged to another subgroup, predominantly found in the Far East (Korea, Japan, and northern China), designated as HBV genotype “Ce.” Overall, the mean ± SD nucleotide sequence difference between HBV genotype Cs and HBV genotype Ce was 4.2% ± 0.3%. When HBV genotype Cs and HBV genotype Ce were compared among patients in Hong Kong, HBV genotype Cs was associated with a higher tendency to develop basal core promoter mutations (80% vs. 50%; \( P = .002 \)), a higher prevalence of C at nucleotide 1858 (95% vs. 0%; \( P < .001 \)), and a lower prevalence of precore stop codon mutations (5% vs. 50%; \( P = .002 \)).

Conclusions. HBV genotype C can be differentiated into 2 subgroups—namely, genotype Ce and genotype Cs—that have different epidemiological distributions and virological characteristics.

Hepatitis B virus (HBV) is divided into 8 different genotypes (A–H) according to the homogeneity of the viral sequence, which is based on an intergenotype divergence of >8% in the complete nucleotide sequence [1]. The prevalences of different HBV genotypes vary geographically and are strongly associated with ethnicity [2–4]. In Asia, HBV genotype C is associated with delayed hepatitis B e antigen (HBeAg) seroconversion [5], more-active hepatitis [6–8], more-advanced liver disease [9], and a higher risk of hepatocellular carcinoma [10, 11], compared with HBV genotype B.

Subgroups have been identified in different HBV genotypes, on the basis of >4% (but <8%) difference in the complete nucleotide sequence. Two subgroups have been identified in HBV genotype A; one is prevalent in Europe (Ae), and the other is prevalent in Africa and Asia (Aa) [12]. Similarly, 2 subgroups have been identified in HBV genotype B [13]; one is found almost exclusively in Japan (Bj), and the other is found in the rest of Asia (Ba). These HBV subgroups may behave differently in terms of prevalence of HBeAg, risk of hepatocellular carcinoma, and response to antiviral treatment [14–16].

Recently, HBV genotype C has been preliminarily classified into at least 2 subgroups, on the basis of phylogenetic analysis of the HBV pre-S gene [17]. So far, no detailed analysis of the full-genome nucleotide sequence of HBV genotype C subgroups has been performed. In the present study, we investigated the characteristics of HBV genotype C subgroups in Hong Kong and their relationship with HBV genotype C subgroups in other parts of Asia.
PATIENTS AND METHODS

Source of patients. Serum samples from 49 patients infected with HBV genotype C, as determined by previous genotype-specific restriction fragment-length polymorphism analysis, were studied [11]. All serum samples were kept in a -80°C freezer for storage. All patients were ethnic Chinese and were followed up in the Hepatitis Clinic of the Prince of Wales Hospital. All patients were positive for hepatitis B surface antigen for at least 6 months and had no evidence of hepatocellular carcinoma. Sixty-nine full-genome nucleotide sequences of HBV genotype C and 12 full-genome nucleotide sequences of non-genotype C HBV were retrieved from the GenBank database for comparison. All reference sequences from GenBank were derived from patients with chronic hepatitis B; HBV nucleotide sequences from patients with acute hepatitis B hepatocellular carcinoma or patients treated with antiviral agents were excluded. The geographical origins of patients harboring different HBV genotype C genomes in GenBank were retrieved from the respective original publications and descriptions in the GenBank database.

Determination of the full-genome nucleotide sequence of HBV. HBV DNA was extracted from 100 μL of serum by use of QIAamp DNA Blood Mini Kit (QIAGEN), in accordance with the manufacturer’s instructions. To obtain the full-length HBV DNA sequence, we performed a seminested polymerase chain reaction (PCR) to amplify 3 overlapping fragments of the HBV genome (A, B, and C) (figure 1 and table 1). For each fragment, 5 μL of the extracted DNA was used with Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Promega) in the first-round PCR and with Taq DNA polymerase alone in the second-round PCR. The final PCR product was examined on a 1.0% agarose/ethidium bromide gel run in 1× Tris-borate/EDTA electrophoresis buffer.

For fragment A, PCR was performed with primers P1 and P2 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 2.5 min), 30 cycles of amplification (94°C for 36 s, 50°C for 36 s, and 72°C for 2.5 min), and a 7-min final extension at 72°C. The PCR product was further amplified in a seminested PCR with primers P1 and P3 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 2 min), 30 cycles of amplification (94°C for 36 s, 52°C for 36 s, and 72°C for 2 min), and a 7-min final extension at 72°C.

For fragment B, PCR was performed with primers P4 and P5 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 30 s), 30 cycles of amplification (94°C for 36 s, 50°C for 36 s, and 72°C for 90 s), and a 7-min final extension at 72°C. The PCR product was further amplified in a seminested PCR with primers P5 and P6 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 90 s), 30 cycles of amplification (94°C for 36 s, 52°C for 36 s, and 72°C for 90 s), and a 7-min final extension at 72°C.

For fragment C, PCR was performed with primers P7 and P9 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 15 s), 30 cycles of amplification (94°C for 36 s, 50°C for 36 s, and 72°C for 2 min 15 s), and a 7-min final extension at 72°C. The PCR product was further amplified in a seminested PCR with primers P8 and P9 in a 5-min initial denaturation at 95°C, followed by 10 cycles of amplification (94°C for 36 s, 60°C for 36 s, and 72°C for 50 s), 30 cycles of amplification (94°C for 36 s, 52°C for 36 s, and 72°C for 90 s), and a 7-min final extension at 72°C.

Figure 1. Location of hepatitis B virus (HBV) genomic fragments and primers for polymerase chain reaction amplification and sequencing of the entire HBV genome.
Table 1. Primers used for polymerase chain reaction (PCR) amplification and sequencing of the hepatitis B virus genome.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence (5′-3′)</th>
<th>Positions, nt</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>TTTTTCACCTCTGCCTAATCA</td>
<td>1821–1841</td>
<td>Sense</td>
</tr>
<tr>
<td>P2</td>
<td>CCCTAGAAAATTGAGAGAAGTC</td>
<td>262–283</td>
<td>Antisense</td>
</tr>
<tr>
<td>P3  *</td>
<td>CCACTGATGCTGGCTAGGATG</td>
<td>3193–3213</td>
<td>Antisense</td>
</tr>
<tr>
<td>P4</td>
<td>GCCGATTCTTGTTGGTACCATA</td>
<td>2801–2824</td>
<td>Sense</td>
</tr>
<tr>
<td>P5</td>
<td>TTCTTTGACATACTTTCCA</td>
<td>979–997</td>
<td>Antisense</td>
</tr>
<tr>
<td>P6  *</td>
<td>TTGGGATGAGCCCTCAGGCTC</td>
<td>3070–3090</td>
<td>Sense</td>
</tr>
<tr>
<td>P7  *</td>
<td>TTGGCCAAAATTCCGAGTC</td>
<td>300–318</td>
<td>Sense</td>
</tr>
<tr>
<td>P8  *</td>
<td>CCCCACCGCTTTGGGGTTCAG</td>
<td>714–734</td>
<td>Sense</td>
</tr>
<tr>
<td>P9  *</td>
<td>GTTGATAAGATAGGGGCGATTTG</td>
<td>2299–2325</td>
<td>Antisense</td>
</tr>
<tr>
<td>Sequencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>CTCCGGAACATTGTTCACCT</td>
<td>2031–2050</td>
<td>Sense</td>
</tr>
<tr>
<td>S2</td>
<td>AAGGTGGGAAAAGTGACTGGG</td>
<td>2469–2490</td>
<td>Sense</td>
</tr>
<tr>
<td>S3</td>
<td>GCTGACGCAACCCCGACTGG</td>
<td>1186–1205</td>
<td>Sense</td>
</tr>
<tr>
<td>S4</td>
<td>TCGATGAGCAAGACCCGGTG</td>
<td>1604–1623</td>
<td>Sense</td>
</tr>
<tr>
<td>S5</td>
<td>GCACACAAAAAGAGAGTACTCA</td>
<td>1940–1959</td>
<td>Antisense</td>
</tr>
<tr>
<td>S6</td>
<td>GTGCCGAATTTCCGAGTC</td>
<td>1441–1460</td>
<td>Antisense</td>
</tr>
<tr>
<td>S7</td>
<td>GAATGAGGGCATAGGGGCGAG</td>
<td>970–991</td>
<td>Antisense</td>
</tr>
<tr>
<td>S8</td>
<td>GAATGAGGGGCAATAGGCGAG</td>
<td>411–433</td>
<td>Antisense</td>
</tr>
<tr>
<td>S9</td>
<td>CATGCTGTAGCTTGGTCC</td>
<td>2831–2850</td>
<td>Antisense</td>
</tr>
</tbody>
</table>

* Primer also used for sequencing.

fication (94°C for 36 s, 52°C for 36 s, and 72°C for 1 min 50 s), and a 7-min final extension at 72°C.

Both strands of PCR products were directly sequenced by use of the Cycling Sequencing Kit DYEEnatic ET Dye Terminator for MegaBACE (Amersham Biosciences). The sequencing products were purified by ethanol precipitation. In each reaction tube, 2 μL of 7.5 mol/L ammonium acetate and 2.5 vol (55 μL) of 100% ethanol was added to a final ethanol concentration of 70% and was centrifuged at 2254 g for 30 min at 14°C. After removal of the supernatant, the DNA pellet was washed again with 100 μL of 70% ethanol and centrifuged at 2254 g for 15 min at 14°C. The resulting DNA pellet was air-dried and resuspended in 10 μL of loading buffer (70% formamide and 1 mmol/L EDTA). Sequences of the amplified HBV DNA were then analyzed using the MegaBACE 1000 DNA sequencer and were aligned into complete genome sequences by use of DNASIS (version 2.5; Hitachi).

Phylogenetic analysis. The full-genome nucleotide sequences of the 49 isolates of HBV genotype C from our center were compared with those of the isolates of HBV genotype C (69 strains) and non–genotype C HBV (12 isolates) retrieved from the GenBank database. The HBV genomic sequences were multiple aligned using CLUSTALW software [18]. The genetic distances were estimated by Kimura’s 2-parameter method, and phylogenetic trees were constructed by the neighbor-joining method [19,20]. The reliability of the phylogenetic tree analysis was assessed and assured by bootstrap resampling with 1000 replicates. Phylogenetic and molecular evolutionary analyses were performed using MEGA (version 3.0) [21]. Subgrouping of HBV genotype C was based on an intersubgroup difference of nucleotide sequence of >4% [22].

Statistical analysis. Differences in nucleotide sequences are expressed as means ± SDs. Proportions were compared by Pearson’s χ² or Fisher’s exact test, as appropriate, and continuous variables were compared by Student’s t test. All tests were 2 sided. Statistical significance was defined as P < .05.

RESULTS

Phylogenetic analysis. In the phylogenetic analysis, we identified 2 major subgroups of HBV genotype C (figure 2A). One major subgroup, designated as HBV genotype “Cs,” included HBV sequences from Southeast Asia (Vietnam, Thailand, Myanmar, Malaysia, and Cambodia) and southern China (south of the Yangtze River). Thirty-nine isolates (80%) from Hong Kong belonged to the HBV genotype Cs subgroup. Of 39 patients with HBV genotype Cs in Hong Kong, 17 had their maternal birthplace traced, all of which were in the Guangdong province of China. Another subgroup, designated as HBV genotype “Ce,” included HBV isolates from the Far East (Korea and Japan) and northern China (north of the Yangtze River). Ten (20%) of the Hong Kong isolates belonged to the HBV genotype Ce subgroup. Five of the 10 patients with HBV genotype Ce in Hong Kong had their maternal birthplace traced; 2 were in
northern China, 2 were in the Guangdong province, and 1 was in the Fujian province. Two outgroups of HBV genotype C were identified from database-derived isolates: 1 from the South Pacific (Micronesia, New Caledonia, and Polynesia) and 1 from the aboriginal Australian population. A similar subtyping could be identified by analysis of the polymerase and X genes (figure 2B and 2C). In the analysis of the large surface gene (pre-S1 to S), the South Pacific HBV outgroup clustered with HBV genotype Ce, since the genomic differences between these 2 HBV genotype C subgroups mainly lay in the polymerase and X regions (figure 2D).

**Intergenotype nucleotide sequence differences.** The nucleotide sequence of HBV genotype Cs differed from that of genotype A by 8.5% ± 0.5%, from that of genotype B by 8.8% ± 0.6%, from that of genotype D by 9.5% ± 0.7%, from that of genotype E by 10.4% ± 0.7%, from that of genotype F by 13.2% ± 0.7%, from that of genotype G by 12.1% ± 0.7%, and from that of genotype H by 13.2% ± 0.7%. The nucleotide sequence of HBV genotype Ce differed from that of genotype A by 8.5% ± 0.5%, from that of genotype B by 8.8% ± 0.6%, from that of genotype D by 9.6% ± 0.6%, from that of genotype E by 10.2% ± 0.6%, from that of genotype F by 13.4% ± 0.7%, from that of genotype G by 12.2% ± 0.7%, and from that of genotype H by 13.5% ± 0.7%. These findings confirmed that both HBV genotype Cs and HBV genotype Ce had >8% nucleotide sequence difference from other HBV genotypes.

**Intersubgroup nucleotide sequence differences.** The intrasubgroup difference in the nucleotide sequences of the entire genome within HBV genotype Cs was 1.9% ± 0.1%, and that for HBV genotype Ce was 2.6% ± 0.2%, whereas the intersubgroup difference between HBV genotype Cs and HBV genotype Ce was 4.2% ± 0.3%. The intersubgroup difference between the nucleotide sequences of the South Pacific outgroup and HBV genotype Cs was 5.2% ± 0.4%, and that between the South Pacific outgroup and HBV genotype Ce was 4.5% ± 0.3%. The intersubgroup difference between the nucleotide sequences of the aboriginal Australian outgroup and HBV genotype Cs was 6.8% ± 0.5%, that between the aboriginal Australian outgroup and HBV genotype Ce was 6.3% ± 0.4%, and that between the aboriginal Australian outgroup and the South Pacific outgroup was 6.2% ± 0.4%. These results supported that HBV genotype Cs and HBV genotype Ce belonged to HBV genotype C and that they could be considered as separate subgroups within this genotype.

**Intrasubgroup nucleotide sequence differences.** Within HBV genotype Cs, 3 clusters of isolates could be identified with at least a 94% bootstrap value (figure 2A). One HBV genotype Cs cluster contained mainly isolates from Thailand and Myanmar, whereas the other 2 HBV genotype Cs clusters contained mainly isolates from Hong Kong, Vietnam, and southern China. The intracluster differences in nucleotide sequences within HBV genotype Cs ranged from 1.9% ± 0.2% to 2.6% ± 0.2%, and these clusters therefore could not be classified as subgroups. Within HBV genotype Ce, 2 clusters of isolates could be identified with an 89% bootstrap value (figure 2A). One HBV genotype Ce cluster included isolates from Okinawa, an island off the southern end of Japan, whereas the other HBV genotype Ce cluster contained the rest of the isolates from the Far East and northern China. The intracluster difference in nucleotide sequences within HBV genotype Ce was 3.5% ± 0.2%.

**Amino acid differences.** Consensus sequences of amino acids were constructed for HBV genotype Cs and HBV genotype Ce by comparing the entire HBV amino acid sequence from the GenBank database and those from Hong Kong in this study. We identified several amino acid sites that differentiated between HBV genotype Cs and HBV genotype Ce in >95% of the sequences (table 2). All of these sites were located in the polymerase region. In a comparison with the reference amino acid sequences of other HBV genotypes, lysine (K) at amino acid 85 in the terminal protein region of the polymerase gene was characteristic of HBV genotype Cs, whereas K at amino acid 143 in the terminal protein region was characteristic of HBV genotype Ce. The polymorphism at amino acid 249 in the spacer region of the polymerase gene was commonly seen among other HBV genotypes.

**Difference in basal core promoter and precore regions.** The differences in basal core promoter and precore regions in the 49 patients from Hong Kong were studied. The ages, sex ratios, HBeAg statuses, alanine aminotransferase levels, and proportions of patients with clinical liver cirrhosis were well matched between patients harboring HBV genotype Cs and those harboring HBV genotype Ce (table 3).

Basal core promoter mutations were detected in 31 (80% [95% confidence interval [CI], 64%–90%]) of 39 HBV genotype Cs isolates and in 5 (50% [95% CI, 19%–81%]) of 10 HBV genotype Ce isolates. The difference fell short of statistical significance (P = .14), but the sample size was small. Twenty-nine patients with HBV genotype Cs and all 5 patients harboring HBV genotype Ce had double mutations (A1762T and G1764A) in the basal core promoter region, whereas 2 patients harboring HBV genotype Cs had a single mutation (G1764A) in the basal core promoter region.

The nucleotide at position 1858 was T in all 10 patients (100%) with HBV genotype Ce but in only 2 patients (5% [95% CI, 0%–12%]) with HBV genotype Cs (P = .001). The remaining 37 patients with HBV genotype Cs had a C at position 1858. The precore stop codon mutation (G1896A) was found only in patients with T1858 and not in patients with C1858. Two patients (5% [95% CI, 0%–12%]) with HBV genotype Cs and 5 patients (50% [95% CI 19%–81%]) with HBV genotype Ce had the precore stop codon mutation (P = .002).
Figure 2. Phylogenetic tree constructed using the nucleotide sequences of 49 hepatitis B virus (HBV) genotype C isolates from Hong Kong and compared with the nucleotide sequences of 69 genotype C and 12 non–genotype C HBV isolates from the GenBank database. Scale bars indicate nucleotide difference. A, Entire genome; B, polymerase gene; C, X gene; D, pre-S to S gene.
Figure 2. (Continued.)
Figure 2. (Continued.)
Figure 2. (Continued.)
On the basis of phylogenetic analysis of the full-genome nucleotide sequence, we have confirmed that there are 2 major subgroups of HBV genotype C in Asia. One subgroup, HBV genotype Cs, is more prevalent in Southeast Asia (Vietnam, Thailand, Myanmar, and southern China), whereas the other subgroup, HBV genotype Ce, is more prevalent in the Far East (Japan, Korea, and northern China). In Hong Kong, 80% of HBV genotype C HBV belongs to the genotype Cs subgroup, and 20% belongs to the genotype Ce subgroup. In addition, these 2 HBV genotype C subgroups have distinct differences in the patterns of mutation in the basal core promoter and precore regions. There are also 2 additional subgroups of HBV genotype C (South Pacific and aboriginal Australian) that appear as outgroups in the phylogenetic analysis.

The majority of the population of Hong Kong is made up of immigrants from the Guangdong province. It is, therefore, not surprising that most patients share the same HBV genotype C subgroup (Cs) as that found in southern China. The minority of patients from Hong Kong who harbor HBV genotype Ce may have ancestors from northern China. Within each HBV genotype C subgroup, several clusters with genomic resemblance to one another can be identified. The most well-defined example is the cluster in Okinawa, where the prevalence of HBV genotype C is much lower than that in the rest of Japan [23]. Since different HBV subgroups are likely to be the results of divergence from genomic mutations over time, knowledge of the geographical distribution and genomic relatedness of the HBV genotype C subgroups will be useful in gaining an understanding of the spread of HBV in Asia.

The prevalence of basal core promoter mutations in HBV genotype C varies widely among different patient cohorts in different geographical regions [10, 11, 17, 21]. In this study, we found that HBV genotype Cs has a higher tendency to have basal core promoter mutations than does HBV genotype Ce. This finding needs to be confirmed in a larger-scale study, because of the small number of cases in the present study. Nevertheless, this may explain the relatively higher prevalence (81%) of basal core promoter mutations in Vietnam, where HBV genotype Cs is the predominant subgroup of HBV genotype C, as well as the lower prevalence (58%) in Japan, where HBV genotype Ce is the predominant subgroup of HBV genotype C [17, 24]. Future studies of HBV genotype C subgroups will be important for gaining an understanding of the prevalence of basal core promoter mutations in other parts of Asia.

Basal core promoter mutations are increasingly recognized as pathogenic in the development of hepatocellular carcinoma [25]. Since basal core promoter mutations are more commonly found in HBV genotype C, the relative importance of basal core promoter mutations and HBV genotypes in the development of hepatocellular carcinoma has been controversial [9, 26–28]. The relationship between basal core promoter mutations and hepatocellular carcinoma is strongest in areas where the background prevalence of these mutations is low, as in Taiwan (29% among HBV genotype C–infected patients without hepatocellular carcinoma) [10], but is weaker in areas where the background prevalence of these mutations is high, as in Hong Kong (67% among HBV genotype C–infected patients without hepatocellular carcinoma) [11]. Since the background prevalence of basal core promoter mutations is likely to be related to the distribution of HBV genotype C subgroups, it would be interesting to know the relative prevalence of HBV genotype Cs and HBV genotype Ce in Taiwan. In the future, the determination of HBV genotype C subgroups may be relevant to the stratification of the risk of hepatocellular carcinoma development.

Nucleotides 1858 and 1896 form a pairing at the hairpin folding of the encapsidation sequence in the precore region [29]. To maintain the stability of the encapsidation sequence, the precore stop codon mutation will not develop in the presence of C1858 [30]. Despite the small sample size in the present study, we have unequivocally identified HBV genotype Ce as being associated with T1858, whereas HBV genotype Cs is associated with C1858. This, in turn, leads to a difference in the prevalence of the precore stop codon mutation between these 2 HBV genotype C subgroups. Patients who are infected with HBV with T1858 tend to develop fewer basal core promoter mutations than those infected with HBV with C1858, possibly because the former group can undergo HBeAg seroconversion by developing the precore stop codon mutation [31].
References


6. Chan HLY, Tsang SWC, Liew CT, et al. Viral genotype and hepatitis B virus DNA levels are correlated with histological liver damage in HBsAg-negative chronic hepatitis B virus infection. Am J Gastroenterol 2002; 97:406–12.


23. Orito E, Ichida T, Sakagawa H, et al. Geographic distribution of hep-


31. Chan HLY, Hussain M, Lok ASF. Different hepatitis B virus genotypes are associated with different mutations in the core promoter and pre-core regions during hepatitis B e antigen seroconversion. Hepatology 1999; 29:976–84.