Genetic diversity of hepatitis B virus in Yunnan, China: identification of novel sub-genotype C17, an inter-genotypic B/I recombinant, and B/C recombinants

Yue Feng,† Jieyu Ran,† Yue-Mei Feng, Jing Miao, Yue Zhao, Yuanyuan Jia, Zheng Li, Wei Yue and Xueshan Xia

Abstract
Yunnan is considered to be a geographical hotspot for the introduction, mutation and recombination of several viruses in China. However, there are limited data regarding the genotypic profiles of hepatitis B virus (HBV) in this region. In this study, we characterized 206 HBV strains isolated from chronic hepatitis B patients in Yunnan, China. Initial genotyping based on 1.5 kb sequences revealed that genotype C was the most prevalent at 52.4 % (108/206), followed by genotype B at 30.6 % (63/206) and unclassified genotypes at 17.0 % (35/206). To characterize the 35 unclassified strains, 32 complete HBV genomes were amplified and analysed; 17 isolates were classified within a known sub-genotype, 8 were classified as B/C recombinants, 1 was classified as a B/I recombinant and 6 constituted a potentially novel C sub-genotype that we designated as C17, based on the characteristics of a monophyletic cluster, >4 % genetic distances, no significant evidence of recombination and no epidemiological link among individuals. Thus, multiple sub-genotypes – namely B1, B2, B4, C1, C2, C3, C4, C8 and C17 – and two distinct inter-genotypic recombinants exist in Yunnan, China, highlighting the complex and diverse distribution pattern of HBV genotypic profiles.

INTRODUCTION
Hepatitis B virus (HBV) infections impact negatively on human health and are a significant healthcare burden worldwide. Chronic HBV infection, which affects approximately 257 million people worldwide [1], is associated with liver damage leading to liver cirrhosis, hepatic decompensation and hepatocellular carcinoma (HCC) [2]. HBV is an enveloped double-stranded (ds) DNA virus that has a relaxed circular DNA genome that is approximately 3200 nucleotides (nt) long and encodes four overlapping open reading frames encoding a polymerase (P), a surface protein (S), a precore/core protein (C) and the X gene product (X) [3, 4]. HBV exhibits high genetic diversity, since the polymerase lacks proofreading activity [5].

To date, eight confirmed (A to H) and two tentative (I and J) HBV genotypes have been identified based on >8 % sequence divergence in the full-length HBV genome [6]. HBV genotypes A, B, C, D, F and I have been further divided into sub-genotypes A1–A7 [7], B1–B9 [8], C1–C16 [9], D1–D10 [10], F1–F6 [11], and I1 and I2 [12] based on 4–8 % nucleotide differences across the genome [5]. HBV recombination is also common; based on the HBV full-length genome, over 400 HBV inter-genotypic recombinants have been characterized and multiple circulation patterns have been generated [13].

In China, nearly 80 million people are chronic carriers, and about 300,000 die from liver cirrhosis and HCC caused by chronic HBV infection each year [14, 15]. Many studies

Received 13 April 2018; Accepted 21 August 2018

Author affiliations: 1Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, PR China; 2Research Institute of Nutrition and Food Science, Kunming Medical University, Kunming, Yunnan 650500, PR China; 3Department of Clinical Laboratory, The First People’s Hospital of Yunnan Province, Kunming, Yunnan, PR China; 4Department of Infectious Disease, The First People’s Hospital of Yunnan Province, Kunming, Yunnan, PR China.

*Correspondence: Yue Feng, fyky2005@163.com; Xueshan Xia, oliverxia2000@aliyun.com

Keywords: chronic hepatitis B; hepatitis B virus; genotype and sub-genotype; genetic diversity; Yunnan.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; dsDNA, double-stranded DNA; GTR, general time-reversible; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; jpHMM, jumping profile hidden Markov model; ORFs, overlapping open reading frame.

†These authors contributed equally to this work.

All sequences obtained in this study have been deposited in GenBank under accession numbers MG826122–MG826153 for 32 complete genome sequences of HBV strains, MG826154–MG826359 for 206 partial P gene sequences of HBV isolates, and MH746811–MH746813 for complete genome sequences of HBV B/I recombinant strain YNKM91.

One supplementary table and two supplementary figures are available with the online version of this article.
have revealed that different HBV genotypes/subgenotypes affect clinical outcome, disease prognosis and response to antiviral treatment in infected patients [16, 17]. HBV genotype B is predominant in southern China, while genotype C is more common in the northern and central areas of the country [18]. Yunnan is located in the most southwestern province of China and shares borders with Myanmar in the west, Laos in the south and Vietnam in the southeast. Several previous reports showed that genotype C is the most abundant, followed by genotypes B, D and I [12, 19]; among genotype B and C isolates, subgenotypes B2 and C2 are predominant in Yunnan, China [20, 21]. However, there is still limited information regarding the diversity and circulation of HBV genotypes and subgenotypes in Yunnan. This was investigated in the present study by examining 206 patients with chronic hepatitis B in Yunnan, China.

RESULTS

Demographic characteristics

The epidemiological characteristics of the 206 chronic hepatitis B patients included in the present study are summarized in Fig. 1. The ratio of males to females was 108 : 98; the age range was 18 to 70 years and the mean age was 37.13 ±12.54 years. Of the 206 participants, 51.0 % (105) were Han, 25.2 % (52) were Yi and 23.8 % (49) were Dai Chinese. The clinical characteristics were as follows: mean ALT, 82.26±98.3 IU l⁻¹; mean AST, 61.65±68.08 IU l⁻¹; mean

![Fig. 1. Maps of the study region and geographical distribution of subjects from all three prefectures of the Yunnan province of southwestern China. The Yunnan province of southwestern China is marked in dark grey, and Kunming city prefecture, Chuxiong prefecture and Dehong prefecture are highlighted in light grey. MYA, Myanmar; THA, Thailand; LAO, Laos; VIE, Vietnam.](image-url)
total protein, 73.81±7.25 IU l⁻¹; mean albumin, 41.66 ±6.72 IU l⁻¹; mean total bilirubin, 16.75±13.79 IU l⁻¹; and HBV DNA, 6.69±1.35 log copies ml⁻¹.

**Phylogenetic analysis of partial HBV genome sequences**

The ~1.5 kb fragments corresponding to HBV partial P, preS1/preS2/S and X genes were amplified and sequenced from 206 HBV-positive subjects with a success rate of 100 %. The results showed that the predominant HBV genotype was C (52.4 %, 108/206), followed by B (30.6 %, 63/206) and unclassified (17.0 %, 35/206) genotypes (Fig. 2a). Phylogenetic analysis revealed that HBV genotype C could be grouped into the distinct subgenotypes C1 (n=60) and C2 (n=48) (Fig. 2c, d), while HBV genotype B comprised subgenotypes B2 (n=61), B1 (n=1) and B4 (n=1) (Fig. 2b). The 35 unclassified HBV isolates formed several distinct monophyletic branches that are distantly related to all known HBV subgenotypes (Fig. 2e), indicating that they represent novel subgenotypes or recombinants.

**Phylogenetic and recombinant analyses based on full-length HBV genome sequences**

To further characterize the 35 unclassified HBV strains, 32 full-length HBV genome sequences were amplified, sequenced, and analysed. The phylogenetic analysis revealed that 16 HBV strains belonged to 5 distinct subgenotypes; 10 were classified within subgenotype C8, 2 each were classified within C4 and C1, 1 was classified within C3 and 1 was classified within C2 (Fig. 3a). The 16 unclassified strains were evaluated for potential recombination events using the jPHMM and SimPlot programs. Nine were identified as potential recombinants. Two of the isolates (YNKM40H and YNKM19H) harboured a recombined form of genotypes B and C (Fig. 3b, c), with a B fragment inserted into the X region (nt 1531–1781) of the C backbone (Fig. 4). Six isolates were identified as genotype B/C recombinants (Fig. 3d–i) with different
recombinant fragment lengths and breakpoints (Fig. 4). Notably, strain YNKM91 was classified as a genotype I/B recombinant (Fig. 3j), with a B fragment inserted into the preS1 region (nt 1–708) of the I backbone (Fig. 4).

Of the remaining seven strains that showed no evidence of recombination, strain YNKM21 was phylogenetically distant from the reported HBV C subgenotypes (Figs 3k and 5a), but was closely related to subgenotype C2, with 95.3% nucleotide identity, indicating a quasi-subgenotype C2.

Interestingly, the other six strains (YNKM10, YNKM31H, YNKM36, YNKM40, YNKM57 and YNKM106) clustered separately from all known C subgenotypes and formed a novel separate cluster with a 90% bootstrap value, suggesting a potentially novel C subgenotype (Fig. 3a).

**Identification and characterization of a novel HBV subgenotype**

To confirm that the six strains belong to a novel C subgenotype, their complete genome sequences were analysed...
phylogenetically, along with HBV reference sequences of representative subgenotypes C1–C16. The results showed that the strains formed a distinct monophyletic cluster supported by a high bootstrap value of 100 % and were distantly related to all known HBV C subgenotypes (Fig. 5b). We therefore tentatively designated these strains as constituting the novel subgenotype C17.

The intergroup nucleotide divergence (mean±SD)% over the full-length genome sequences of C17 was compared to that of representative subgenotypes (C1–C16) (Fig. 5b). The result revealed that C17 were different from known HBV subgenotypes of C1–C16 by 4.1–6.1 %. An amino acid sequence comparison between HBV C17 isolates and other subgenotypes (C1–C16) indicated that C17 had 17 unique amino acid motifs within the P gene and preS1/S2 and X regions (Fig. 5c).

In addition, six HBV C17 strains in this study were analysed for mutations in the reverse transcriptase (RT) domain of the polymerase gene, the preS1, preS2 and S genes, the basal core promoter (BCP) and the precore (PC) gene that might be associated with resistance to therapy and disease progression (Table S1, available in the online version of this article). The previously reported putative nucleos(t)ide analogue resistance (NAr) mutations at I91L, W153Q and I233V were identified in 33.3 % (2/6), 16.7 % (1/6) and 16.7 %, respectively. The pretreatment substitutions at Y124H and I224V were found in 33.3 % (2/6) and 100 % (6/6), respectively. Notably, the C17 strain possessed several mutations reported to be associated with HCC, including R24K (50 %, 3/6) and L126T/S (33.3 %, 2/6) in the S protein, T1753C (83.3 %, 5/6) and G1764A (100 %, 6/6) in the BCP region, and G1896A (100 %, 6/6), A2159G (33.3 %, 2/6) and (50 %, 3/6) in the PC gene.

**Distribution of HBV genotypes, subgenotypes, and recombinants**

The HBV genotype, subgenotype and recombinant distributions among the 206 patients with chronic Hepatitis B in Yunnan, China are summarized in Fig. 6. Notably, in the sampled prefectures, subgenotype C2 was more prevalent in Kunming than in Dehong (39/105, 37.1 % vs 0/96, 0 %).
However, the prevalence of subgenotype C1 was higher in Dehong (33/49, 67.4%) than in Kunming (7/105, 6.7%) or Chuxiong (22/52, 42.3%) \((P<0.05)\). The HBV strains circulating in Kunming exhibited much higher genotypic diversity than those in the other two prefectures (Fig. 6).

**DISCUSSION**

In the present study, we evaluated the HBV genotype, subgenotype and recombinant distributions among 206 patients with chronic hepatitis B in Yunnan, China. Our results showed that genotype C is the predominant genotype (63.1%) in Yunnan followed by B (30.6%), which is consistent with previous findings (76.9 and 15.4%, respectively, in Yunnan [22] and 62.5 and 33.3%, respectively, in Kunming [19]). We identified multiple subgenotypes (B1, B2, B4, C1, C2, C3, C4, C8 and C17), two intergenotypic recombinants (B/C and B/I) and a quasi-subgenotype (C2). Our results demonstrate that Yunnan has a high degree of HBV genotype diversity and complexity, posing a challenge for antiretroviral therapy to control HBV infection.

The C1 subgenotype of HBV was predominant in subjects from the Dai ethnic group from Dehong and was more prevalent there than in Kunming and Chuxiong. This may be explained by the geographical location of the Dehong prefecture, which shares a border with Myanmar, where there is a very high prevalence of the HBV C1 subgenotype [22]. It has also been reported that HBV C1 is only found in Southeast Asian countries, including Vietnam, Myanmar and Thailand [23]. We therefore speculate that the HBV subgenotype C1 circulating in Dehong originated in Myanmar. It is also worth noting that a high genetic diversity of HBV, including eight HBV subgenotypes and two different recombinant forms, was observed among individuals from the Han ethnic group in Kunming harbouring HIV-1. This may be due to the fact that Kunming is a provincial capital city and also a centre of tourism and trade with other provinces in China and Southeast Asia.

**Fig. 5.** Analysis of full-length genome sequences of novel HBV subgenotype C17. (a) Recombinant analysis was performed using Simplot 3.5.1 software based on 100 replicates with a 300 bp sliding window moving in steps of 50 bases. (b) Maximum-likelihood tree based on phylogenetic analysis of the C subgenotypes (C1–C17) and the estimated intergroup nucleotide divergence between C17 and the other known C subgenotypes (C1–C16). (c) Signature pattern analyses of the newly named HBV subgenotype, C17 \(^{b}\) based on the complete genome.
Asia. The frequent contact among people from different regions may have resulted in multiple introductions of alien HBV strains, which then mixed with indigenous ones. This is supported by our previous finding that the clusters of subgenotypes B2 and C2 comprise strains from China and other East Asian countries, while subgenotypes C1, C5 and C7 and genotype I form a cluster with strains from Southeast Asia [19].

The most important finding of this study was the identification of a novel HBV subgenotype. The characteristics of a highly supported monophyletic cluster, >4% genetic distance, no significant evidence of recombination and unique amino acid signatures, and the fact that the six strains in question were isolated from individuals with no epidemiological link, are consistent with the standard classification of a novel HBV subgenotype [24, 25], which we designated as C17. In terms of nucleotide sequence similarity and phylogeny, C17 was most closely related to HBV subgenotype C8, suggesting that they originate from the same HBV lineage.

Intergenotype recombination plays a key role in HBV emergence and genetic variability and presents enormous challenges for vaccine design and antiviral therapy strategies [26]. We found a relatively low frequency of intergenotypic recombinant HBV (4.4%), of which most strains were genotype B/C recombinants. Notably, these isolates showed a variety of recombination breakpoints, including at nt 400–667, 770–1458, 1500–1800 and 2771–3200 (Fig. 4), suggesting that there is no hotspot for recombination. Another important finding of our study is the first observation of a B/I intergenotype HBV recombinant. Previous studies have shown that intergenotypes exhibit circulation patterns that are similar to those of their original
genotypes [13]; for instance, B/C recombinants are limited to South and East Asia and are most prevalent in regions where HBV genotypes B and C are widespread, including China, Indonesia, Japan, Malaysia, Philippines, Taiwan, Thailand and Vietnam [27, 28]. It is therefore tempting to speculate that the B/I recombinant arose from the abundance of genotypes B and I circulating in Yunnan. In fact, while Yunnan has a high prevalence of genotype B, genotype I is uncommon. Relatively few strains causing HBV genotype I infection have been reported in Yunnan [12, 19]. We did not observe any genotype I strains in the present study, possibly due to the small study population. HBV intergenotype recombination is easily detected on the basis of phylogenetic, simplot and bootscanning analyses. However, mixed infection by two different HBV genotypes or subgenotypes can result in an erroneous conclusion that recombination has occurred. This may be attributed to the fact that recombination analysis performed on PCR products can be biased by template switching in the first rounds of PCR amplification, or the splicing of the full-length sequences using subgenomic PCR. To eliminate this, it is necessary to select a genotype-independent PCR amplification strategy and obtain complete HBV genome sequences from a unique PCR fragment [13]. In this study, to demonstrate that the strain YNKM91 is a B/I intergenotype HBV recombinant, we obtained three full-length genomes from a unique PCR fragment and repeated the PCR three times on separate occasions (MH746811–MH746813). Our analyses of nucleotide similarities and phylogenetic (Fig. S1a), recombinant (Fig. S1b) and subregion trees (Fig. S2) confirmed that the strain YNKM91 really belongs to a B/I recombinant, not an artifact.

Our study documents the complex and diverse distribution patterns of HBV subgenotypes and intergenotype recombinants in Yunnan, China. We identified a novel HBV subgenotype (C17) and characterized a previously unreported B/I intergenotype recombinant. These findings provide an insight into the genetic characteristics of HBV strains in Yunnan, which can be used to develop effective measures to prevent the transmission of HBV in this region.

METHODS

Study population

Plasma samples were collected between June 2016 and September 2017 from 206 chronic hepatitis B patients from 3 ethnic groups (Han, Yi and Dai) residing in Kunming, Chuxiong and Dehong prefectures, respectively, in Yunnan, China. Kunming is the capital and largest city of Yunnan province, southwest China and is a multi-ethnic city; the majority of the population are Han and Yi, followed by Hui, Bai, Miao, Hani, Zhuang, Dai, Lisu and other ethnicities. Chuxiong is located in the central part of Yunnan Province and abuts Kunming to the east. The residents are mainly made up of the Yi ethnic minority (Yi autonomous prefecture). Dehong is situated at the western boundary of Yunnan province and shares a boundary with Myanmar; 51.83% of the population comprise minority ethnicities, mostly Dai and Jingpo.

The retrieval location and quantity of the samples collected for this study are shown in Fig. 1. The samples met the following inclusion criteria: (1) hepatitis B surface antigen-positive for 6 months with normal serum alanine aminotransferase (ALT) levels; (2) subject was residing in Yunnan province and was over 18 years old; (3) complete demographic information and clinical data were available for subject; (4) subject consented to the use of patient information in studies on HBV epidemics; and (5) subject was treatment-naïve during sampling. Demographic information and data on risk factors for HBV transmission were collected through a questionnaire. Standard clinical biochemical tests for hepatic and renal function and blood glucose and fat were performed at the time of sampling. The remaining plasma samples were aliquoted and stored at −80 °C. Demographic information relating to age, gender and ethnicity was obtained via self-report questionnaires. The clinical parameters of disease progression, including ALT and aspartate aminotransferase (AST) levels and HBV DNA load, were determined at the time of sampling. Plasma was separated from whole blood samples using EDTA tripotassium salt and stored at −80 °C for HBV DNA extraction. The study protocol was approved by the First People’s Hospital of Yunnan Province Ethics Committee (YNFPH-2016–012). Written informed consent was obtained from all participants.

DNA extraction and HBV gene amplification

HBV DNA was isolated from 200 µl plasma using the AxyPrep Blood Genomic DNA Miniprep kit (Axygen Biosciences, Union City, CA, USA) according to the procedure described in the manual. To determine the HBV genotype and subgenotype, partial sequences of nearly 1.5 kb (GQ358158 : 55–1536) were amplified using the nested PCR strategy with two oligonucleotide primer pairs. HBVspF01 : 5’-CTGCTGGTGGCATGGAG-3’ and HBVspR01 : 5’-GGCGTCCAGTTCAGGRAYAGT-3 and HBVspF01 : 5’-CTGCTGGTGGCATGGAG-3’ and HBVspR01 : 5’-GGCGTCCAGTTCAGGRAYAGT-3 were used for the second round of the following cycling conditions: 94°C for 5 min; 94°C for 30 s, 55°C for 30 s, 72°C for 2 min 30 s, 30 cycles; 72°C for 10 min. HBVspF02 : 5’-GGRAYAstrAACCCTGT TTCACGCTGGTGTCCATGCAGG-3’ were used for the first round with the following cycling conditions: 94°C for 5 min; 94°C for 30 s, 55°C for 30 s, 72°C for 2 min 30 s, 30 cycles; 72°C for 10 min. To confirm the novel subgenotype and recombinant, and to eliminate the possibility of mixed infection by two or more different HBV genotypes in particular, complete HBV sequences from a unique PCR fragment were amplified as previously described [29, 30]. Amplified PCR products were detected by electrophoresis on an 1.0 % agarose gel under ultraviolet illumination and purified using a DNA purification kit (Tiangen, Beijing, People’s Republic of China). The products were sequenced by Shenzhen Invitrogen Biotechnology Co. (ShenZhen, People’s Republic of China) on an ABI
Phylogenetic analysis

The sequencing data were initially verified using the National Center for Biotechnology Information’s Basic Local Alignment Search Tool (BLAST). Reference sequences in GenBank were used for a comparative analysis of all HBV genomic sequences. Multiple alignments of the selected sequences were performed using Clustal v.1.8.1 software. Phylogenetic trees were constructed based on the obtained datasets with the maximum-likelihood v.7.1.5 software. Phylogenetic trees were constructed based on the maximum-likelihood method using MEGA v.6.0.6 and the general time-reversible plus gamma distribution plus invariant sites (GTR+G+I) model. Bootstrap values were calculated based on 1000 replications of the alignment.

Recombination analysis

Potential HBV genetic recombination events were analysed using the jumping profile hidden Markov model (pHMM) included in the Los Alamos human immunodeficiency virus (HIV) database, and were further validated with SimPlot v.3.5.1 software to generate similarity plots and perform bootscanning and informative site analyses. The structure of the new HBV recombinant forms was elucidated based on information on breakpoints obtained from these analyses.

Funding information

This study was funded by the National Natural Science Foundation of China (81601767), the Key Project of Yunnan Education Department (2015Z073), New Product Development Projects of Yunnan Province (2016BC005) and Yunnan Provincial Innovation Team Project (2015HCO30).

Conflicts of interest

The authors declare that there are no conflicts of interest.

References


