scientific reports

Check for updates

OPEN

Distinct phylogeographic distributions and frequencies of precore and basal core promoter mutations between HBV subgenotype C1 rt269L and rt269I types

Dong Hyun Kim[1,2], Yu-min Choi[1,2], Junghwa Jang[1,2,3], Ziyun Kim[1,2,3] & Bum-Joon Kim[1,2,3,4]✉

Hepatitis B virus (HBV) genotype C exhibits two distinct polymorphisms in its viral polymerase: rt269I and rt269L. Recently, we reported that there are distinct virological and clinical profiles between chronic patients with subgenotype C2 with the rt269I polymorphism and those with the rt269L polymorphism, with the latter being more closely related to liver disease severity. This study explored the phylogenetic and geographic distributions, as well as the mutation frequencies, of precore (T1858C/G1896A) and basal core promoter (BCP) (A1762T/G1764A) mutations between these two types within the HBV subgenotype C1. Analysis of 408 HBV/C1 full-genome sequences from GenBank revealed clear phylogenetic separation between rt269L and rt269I in subgenotype C1. Geographically, rt269I strains within subgenotype C1 are predominant in Southwest Asia (e.g., Thailand and Bangladesh), whereas rt269L strains are more common in East Asia and Southeast Asia (e.g., Vietnam, China, and Hong Kong). Notably, compared with rt269L in subgenotype C2, rt269I presented higher frequencies of the C1858 and BCP mutations but lower frequencies of the G1896A mutation. These findings suggest significantly distinct phylogeographic and mutational characteristics of the rt269L and rt269I types of subgenotype C1, impacting clinical outcomes and evolutionary trajectories.

Keywords Hepatitis B virus (HBV), Genotype C, Sub-genotype C1, rt269I type, r269L Type

Hepatitis B virus (HBV) is a major global health problem that affects over 240 million people worldwide and causes approximately 786,000 deaths annually due to severe liver diseases such as chronic hepatitis, liver failure, cirrhosis, and hepatocellular carcinoma (HCC)[1,2]. Chronic HBV infection is responsible for 60% to 80% of HCC cases in Asian patients; the incidence in Asia is notably higher than that in other regions because long-term chronic infection is often acquired through vertical transmission at birth or during early infancy[3].

HBV belongs to the *Hepadnaviridae* family and possesses a partially double-stranded DNA genome[4]. Owing to an error-prone reverse transcription process and a lack of proofreading ability of the viral polymerase, the mutation rate of HBV is significantly higher than that of other DNA viruses, leading to considerable genetic diversity within the virus[5,6]. This genetic diversity is manifested in at least 10 genotypes and more than 30 subgenotypes[7]. Among the known genotypes, HBV genotype C (HBV/C) is predominantly found in Asia, one of the regions with the highest burden of chronic HBV infection[8,9]. Compared with genotype B, genotype C is associated with more severe liver diseases, delayed HBeAg seroconversion, and a greater risk of developing HCC, which is also common in Asia[10,11]. Additionally, individuals with genotype C infections often have a poor response to interferon therapy[6].

[1]Department of Microbiology and Immunology, College of Medicine, Seoul National University, Seoul, Republic of Korea. [2]Department of Biomedical Sciences, College of Medicine, Seoul National University, Seoul, Republic of Korea. [3]BK21 FOUR Biomedical Science Project, Seoul National University College of Medicine, Seoul, Republic of Korea. [4]Institute of Endemic Disease, Seoul National University Medical Research Center (SNUMRC), Seoul, Republic of Korea. ✉email: kbumjoon@snu.ac.kr

HBV genotype C has been further categorized into two distinct subgenotypes, referred to as Cs (Southeast) and Ce (East). These subgenotypes have been designated as C1 and C2, respectively. In Asia, where infections with HBV genotype C are predominantly endemic, the majority of cases are associated with subgenotypes C1 and C2[12]. Subgenotype C1 is found mainly in Southeast Asia, including countries such as Vietnam, Thailand, Myanmar, Malaysia, and Cambodia, whereas subgenotype C2 is prevalent in East Asia, including Korea, Japan, and China[13]. Generally, subgenotype C2 leads to more aggressive liver diseases than subgenotype C1 does, indicating significant clinical differences between these subgenotypes[14].

Recent studies have highlighted the distinct clinical outcomes associated with the rt269L and rt269I mutations in HBV genotype C2[15–19]. In this study, we conducted an analysis of the mutation frequencies and phylogenetic distribution of rt269L and rt269I using HBV genotype C genome sequences retrieved from public databases. Subsequently, we compared the sequence patterns of the rt269 variants in HBV genotype C patients. Additionally, we investigated the geographical distribution of subgenotype C1 with respect to the rt269 variants and assessed the frequency of basal core promoter and precore mutations associated with each subgenotype. Furthermore, we aimed to elucidate how these polymorphisms influence the prevalence of clinically relevant mutations and their relation to geographical dynamics.

Materials and methods

Patient serum collection

Serum samples were obtained from 320 patients who attended Konkuk University Hospital between 2013 and 2015. Patients were included in the study based on the criteria of hepatitis B surface antigen (HBsAg) positivity and HBV DNA positivity for a duration exceeding six months. All patients were treatment-naïve for lamivudine, adefovir dipivoxil, entecavir, telbivudine, tumor necrosis factor inhibitors, and pegylated interferon (peg-IFN). HBV DNA was extracted from 200 µL of serum using the QIAamp DNA Blood Mini Kit (QIAGEN Inc., Hilden, Germany). All experimental protocols received approval from the Institutional Review Board of Konkuk University Hospital (KUH-1010544) and Seoul National University Hospital (IRB1012-131–346). The protocols were conducted following the guidelines and regulations set by the Institutional Review Board. The experiments were carried out using virion DNA extracted from isolates; therefore, informed consent was not required, and a waiver of informed consent was approved by the institutional review boards (IRBs). Following the method described in a previous study[20], typing at the rt269 position in the HBV genome from patient serum was conducted, identifying 22 patients with co-dominant rt269L and rt269I variants **(Supplementary Table 1)**. Among these, serum DNA from patients B182 and B184 were selected and submitted for further analysis.

Nested PCR Amplification

A nested PCR-based sequencing protocol was employed to analyze mutation patterns and frequencies in the RT region. PCR amplification was conducted using the same primers and cycles as those described in our previous study[21]. The first round of PCR was conducted with the sense primer Pol-RT-F1 (5'-CAGCCTACTCCCATCTCT CCACCTCTAAG-3') and the antisense primer Pol-RT-R1 (5'-GCTCCAGACCGGCTGCGAGC-3'), producing a 1375-bp fragment between positions 3157 nt and 1316 nt of the HBV genome In the second round, the sense primer Pol-RT-F2 (5'-CCTCAGGCCATGCAGTGGAA-3') and the antisense primer Pol-RT-R2 (5'-GTATGGA TCGGCAGAGGAGC-3') were used allowing for approximately 100 bp flanking each side of the 1032 bp reverse transcriptase region. This approach generated a 1291 bp product spanning positions 3196 nt to 1271 nt of the HBV genome. The thermal cycling conditions for both PCR rounds consisted of an initial denaturation at 95 °C for 10 min, followed by 15 cycles of 94 °C for 15 s, 55 °C for 15 s, and 68 °C for 3 min, with a final extension at 72 °C for 5 min. In the second round, 2 µL of the first-round product was used, with the same conditions applied, except for 30 cycles.

TOPO Cloning-based Sequencing

For clone-based sequencing, PCR products (0.5 to 4 µl) were mixed with salt solution, TOPO® vector, and water to a final volume of 6 µl. The mixture was incubated at room temperature for 5 min and then incubated on ice for transformation. Competent cells (30 µl) were added to 3 µl of the reaction mixture, incubated on ice for 2–3 min, heat-shocked at 42 °C for 45 s, and returned to ice for 5 min. SOC medium (125 µl) was added, and the mixture was shaken at 37 °C for 1 h. Then, 100 µl of competent cells and 250 µl of SOC were spread onto LB plates and incubated overnight at 37 °C. The next day, 10 colonies from each plate were selected for colony PCR and sequencing. The assembled sequences are registered in GenBank (accession nos. PQ008220 to PQ008234).

Genome database construction

The HBV genome sequences, excluding Genotype C, utilized a dataset of 1351 sequences from previous studies[22]. For Genotype C, additional sequences were obtained from the NCBI GenBank by searching for "HBV," "complete genome," and "genotype C" (2023.06). Sequences that overlapped with the preexisting database were excluded. The HBV subgenotype C1 dataset comprised a total of 408 sequences. Subgenotype C2 sequences were sourced from a database of 683 sequences used in a previous study[23]. A total of 119 HBV genotype C (C4, C5, C8, C10, C11) genomes other than subgenotypes C1 and C2 were included. According to a previous study in the database, subgenotype C3 cannot be clearly separated from subgenotype C1[22]. In addition, subgenotypes C6, C7, and C9 were excluded from the database construction for this study because of the small number of sequences and lack of clear separation. All sequences used in this study are listed in **Supplementary Table 2**.

Phylogenetic analysis

Using the genomic sequence of HBV genotype C from public database, alignment was performed by applying MAFFT with the FFT-NS-2 algorithm in Geneious Prime software (2024.0.2, Biomatters, Inc., New Zealand)[24],

and all indels were removed to align with the complete genome of HBV genotype C (3215 bp). HBV reverse transcriptase sequences extracted from serum were aligned with subgenotype C2 HBV sequences[25]. Phylogenetic trees were constructed using FastTree software via the maximum likelihood method[26]. The tree was annotated using iTOL web-based software with additional data, codon typing data for the rt269 region and geographical location data[27].

Genetic diversity analysis

Genetic diversity was calculated by generating consensus sequences for each HBV subgenotype using Geneious Prime software (2022.1.1, Biomatters, Inc., New Zealand)[16]. A consensus sequence was generated from the alignments for each rt269L and rt269I type. This consensus sequence was aligned with each type that occurred. Pairwise distances between individual sequences and the consensus sequence within each subgenotype were then calculated using MEGA X (Tamura et al., 2021)[28].

Mutation and signature sequence analysis

In the analysis of the signature sequence of each type, consensus was defined as the nucleotide being shared by more than 50% of strains at each position of the sequence according to the majority rule, and the consensus sequences for rt269L and rt269I were extracted from the alignment of each type. The consensus sequences of each extracted type were aligned, and regions where sequences differed from each other were identified and designated as signature sequences. The proportions of nucleotide configurations for each position were calculated and compared between types. The mutation frequencies of the basal core promoter mutation (A1762T, G1764A) and the precore mutation (T1858C, G1896A) were calculated.

Statistical analysis

Statistical analysis was conducted using GraphPad Prism 8.0 (GraphPad Software, San Diego, CA, USA). Mutation sites were compared using the chi-square test, with significance values indicated in the figures and tables. A P value < 0.05 was considered statistically significant.

Results

Phylogenetic distribution of rt269L and rt269I within HBV genotype C

Unlike most other genotypes that exhibit a single type, rt269I, genotype C has two types of polymorphisms, rt269L and rt269I, in the 269th codon of the HBV reverse transcriptase gene (rt269). In all the genotypes except genotype C, isoleucine (rt269I) was present in more than 94% of the sequences (Table 1). However, in genotype C, the frequencies of leucine (rt269L) and isoleucine (rt269I) were 44.5% (539 strains/1210) and 55.0% (665 strains/1210), respectively. First, to investigate the phylogenetic distribution between the rt269L and rt269I types in various subgenotypes within genotype C, phylogenetic analysis using HBV genotype C sequences available from NCBI public databases was performed. We found that of the seven subgenotypes, C1, C2, C4 and C8 had both the rt269L and rt269I types, but the other three subgenotypes, C5, C10 and C11, had the single rt269I type (Table 2). In this study, we focused on the phylogenetic distribution of two subgenotypes, C1 and C2, which represent most genotype C infections (1091strains/1210, 90.2%). With some exceptions, the phylogeny of subgenotype C1 revealed a clear separation between the rt269L and rt269I types, suggesting phylogenetic stability between both types in subgenotype C1 (Fig. 1). In contrast, the phylogeny of subgenotype C2 demonstrated no clear phylogenetic divergence between rt269L and rt269I. Taken together, these results showed that unlike subgenotype C2, where the rt269L and rt269I types are intermixed, subgenotype C1 demonstrated distinct phylogenetic clustering for each type. The genetic diversity analysis revealed that the rt269I type exhibited significantly greater genetic diversity than the rt269L type in subgenotype C1 (C1: $p < 0.0001$) and in subgenotype C2 (C2: $p = 0.0011$) (Fig. 2). However, disparity in the genetic diversity between the types was more pronounced in subgenotype C1 than in subgenotype C2, further supporting clear phylogenetic separation between rt269L and rt269I in subgenotype C1[16]. This finding indicates phylogenetic stability and separation between the rt269L and rt269I types in subgenotype C1.

Genotype	Total no	No. rt269L/I	L % / I %
A	259	13/244	5.0%/94.2%
B	687	20/661	2.9%/96.2%
C	1210	539/665	44.5%/55.0%
D	549	22/527	4.0%/96.0%
E	145	0/145	0%/100%
F	80	3/77	3.8%/96.3%
H	11	0/11	0/100%
Total	2941	597/2330	20.3%/79.2%

Table 1. Distribution of rt269L and rt269I in various HBV genotypes. This table presents the distribution of leucine (rt269L) and isoleucine (rt269I) variants at the rt269 site across various HBV genotypes. For each genotype, the table lists the total number of sequences analyzed (No.), the number of sequences containing rt269L or rt269I (No. rt269L/I), and the percentage distribution of leucine and isoleucine (L% / I%) relative to the total number of sequences in the dataset.

Genotype	Total no	No. rt269L/I	L % / I %
C1	408	197/209	48.3%/51.2%
C2(1)	261	156/102	59.8%/39.1%
C2(2)	270	92/178	34.1%/65.9%
C2(3)	152	86/66	56.6%/43.4%
C4	22	6/15	27.3%/68.2%
C5	18	0/18	0/100%
C8	17	2/15	11.8%/88.2%
C10	22	0/22	0/100%
C11	40	0/40	0/100%
Total	1210	539/665	44.5%/55.0%

Table 2. Distribution of rt269L and rt269I in HBV genotype C. This table presents the distribution of leucine (rt269L) and isoleucine (rt269I) polymorphisms at the rt269 site across HBV genotype C. For each subgenotype, the table lists the total number of sequences analyzed (No.), the number of sequences containing rt269L or rt269I (No. rt269L/I), and the percentage distribution of leucine and isoleucine (L% / I%) relative to the total number of sequences in the dataset.

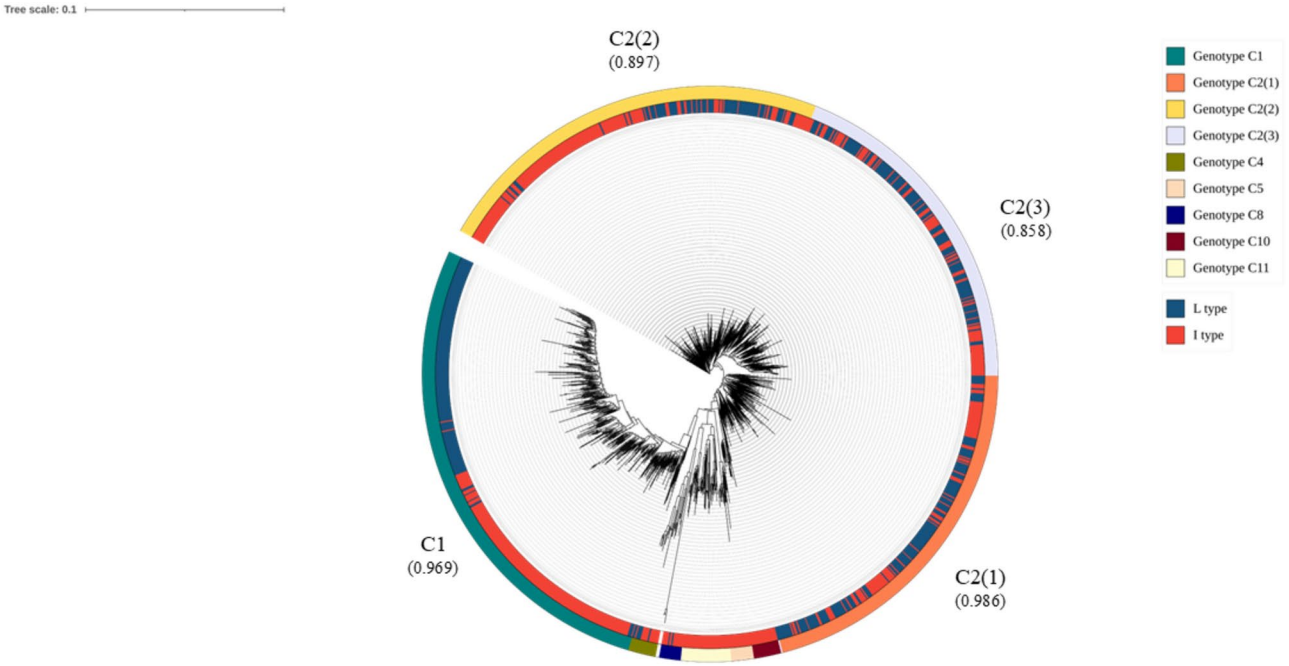

Fig. 1. Phylogenetic distribution of rt269L and rt269I in various subgenotypes of HBV genotype C. This phylogenetic tree was constructed utilizing the whole genome sequences (3215 bp) of 1,402 HBV genotype C with three genotype A outgroups. This figure illustrates the distribution of leucine (L type, blue) and isoleucine (I type, red) at the amino acid of rt269 site across HBV genotype C in the inner ring. The outer ring represents different subgenotypes: C1, C2(1), C2(2), C2(3), C4, C5, C8, C10, and C11, each distinguished by a unique color as indicated in the legend. A FastTree support value is listed beneath the name of each main clade. This illustration demonstrates the prevalence and evolutionary relationships of rt269 variants within and between these HBV genotypes.

Distinct sequence types of the rt269L and rt269I types in subgenotypes C1 and C2

Sequence types in the rt269 codon between subgenotypes C1 and C2 were examined. In subgenotype C1, the TTG sequence type was predominant in the rt269L type, representing 88.3% (174/197) of the sequence types in the rt269 region (Table 3). For the rt269I type, the frequency of the ATC sequence type was 95.7% (200/209). Therefore, two nucleotides differ between the major rt269L (**T**T**G**) and rt269I (**A**T**C**) sequence types in subgenotype C1, possibly acting as a hurdle for quasispecies generation between both types in an infected patient. In contrast, for subgenotype C2, the CTC sequence type was the most common in the rt269L type [88.9% (297/334)], whereas the ATC sequence type was the most predominant for the rt269I type [88.7% (307/346)], similar to subgenotype C1. Therefore, only one nucleotide differs between the major rt269L (**C**TC)

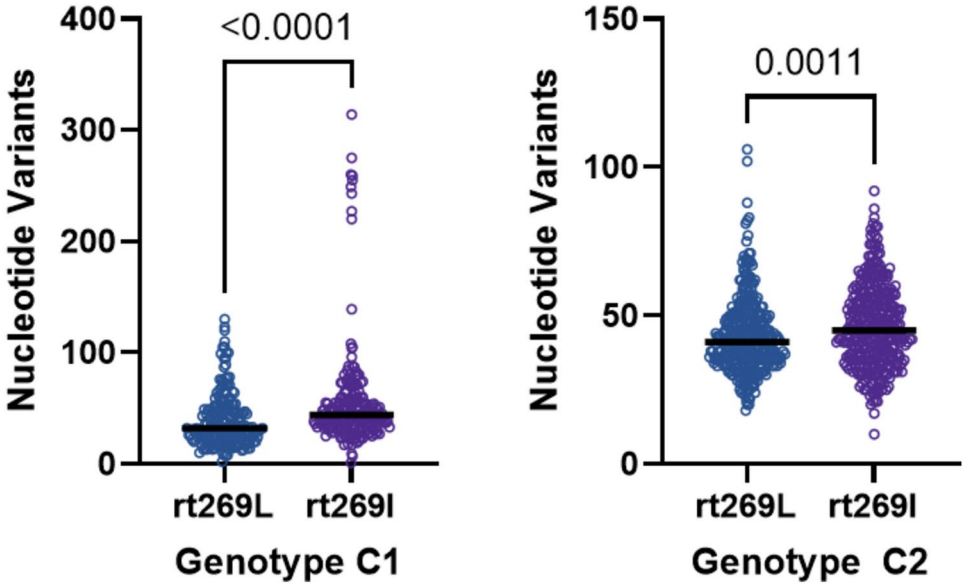

Fig. 2. Genetic diversity comparison of HBV subgenotype C according to nucleotide variants of rt269 Site. Violin plots illustrating the genetic diversity of the rt269L and rt269I variants within HBV sub-genotypes C1 and C2. The p-values indicate statistically significant differences in genetic diversity between the variants (**: $p < 0.01$, ****: $p < 0.0001$).

Genotype	Amino acid	Codon	No		Total
C1	Leucine (TTG, TTA, CTC, CTT)	TTG	174 (88.3%)	197 (48.3%)	408
		TTA	1		
		CTC	14		
		CTG	5		
		CTT	3		
	Isoleucine (ATC, ATT)	ATC	200 (95.7%)	209 (51.2%)	
		ATT	9		
	etc	MTC, GTG	2	2 (0.5%)	
C2	Leucine (TTG, CTC, CTT, CTA)	TTG	0	334 (48.9%)	683
		CTC	297 (88.9%)		
		CTG	1		
		CTT	18		
		CTA	18		
	Isoleucine (ATC, ATT)	ATC	307 (88.7%)	346 (50.7%)	
		ATT	39		
	etc	MTC, TTC	3	3 (0.4%)	

Table 3. Codon distribution at the rt269 site in HBV subgenotypes C1 and C2. This table shows the distribution of amino acids and corresponding codons in the rt269 region across different HBV genotypes. Genotype C sequences from a public database consisted of 408 sequences for subgenotype C1 and 683 sequences for subgenotype C2. For each subgenotype, the specific amino acid (leucine or isoleucine) present, the codon encoding this amino acid, the number of occurrences (counts) and the total number of sequences analyzed for each genotype are listed.

and rt269I types (**A**TC) of subgenotype C2, which could indicate rapid evolutionary changes between the rt269L and rt269I types compared with subgenotype C1. To assess the quasispecies dynamics between the rt269L and rt269I types in subgenotype C2, we evaluated two Korean patients, B182 and B184 (Fig. 3). In total, 8 out of 10 colonies (B182: 4 L type, 4 I type) and 7 out of 10 colonies (B184: 4 L type, 3 I type) were sequenced from each plate. Our data revealed that both types coexisted in each patient and were phylogenetically related to each other, supporting our hypothesis that the distribution between the rt269L and rt269I types of subgenotype C2 is due to viral quasispecies generation rather than phylogenetic evolution.

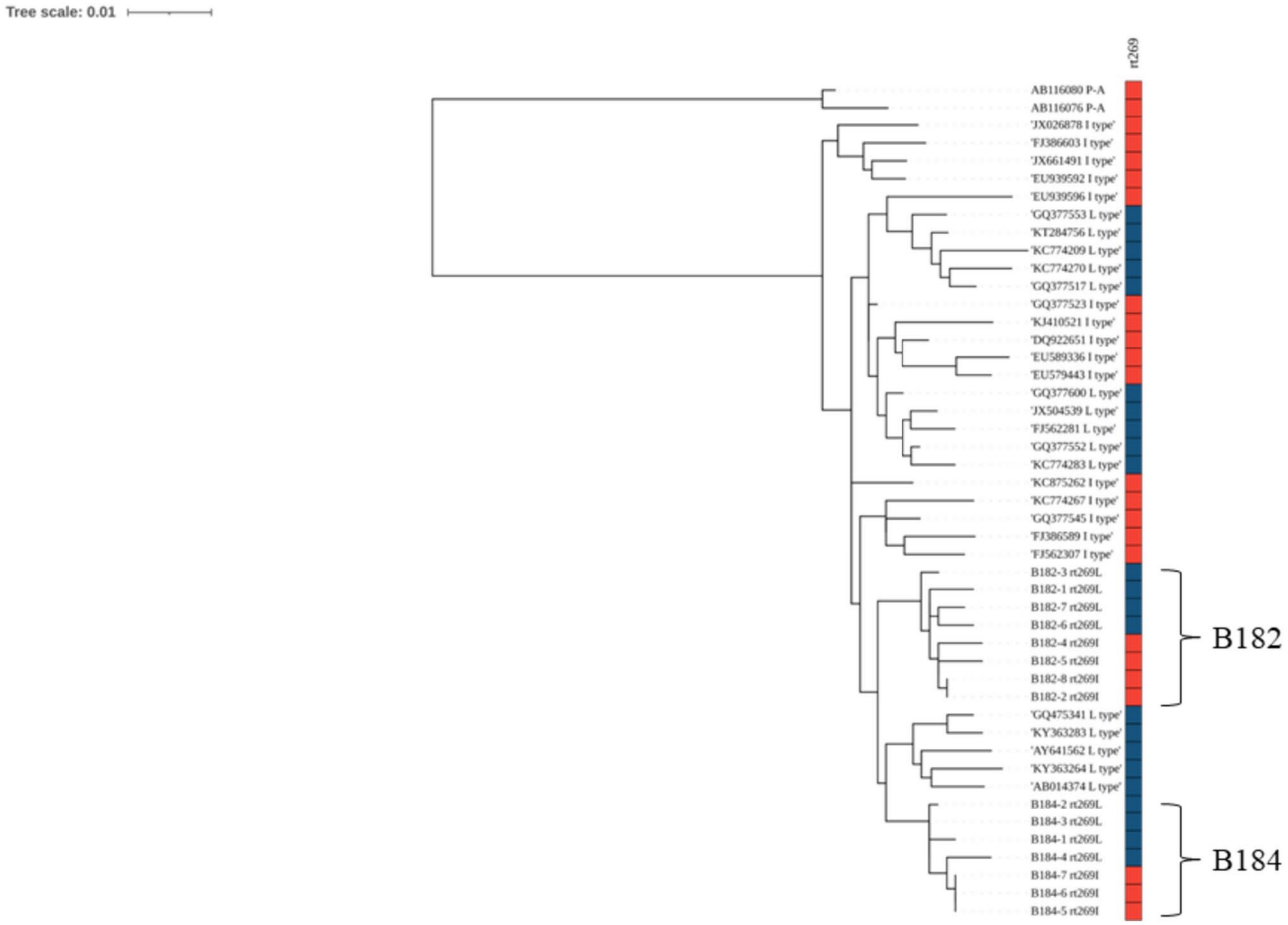

Fig. 3. Phylogenetic distribution of quasispecies of HBV subgenotype C2 in patients. This phylogenetic tree was based on 30 subgenotype C2 and 15 serum-derived reverse transcriptase regions (1032 bp), with three genotype A outgroups. illustrates the distribution of leucine (L type, blue) and isoleucine (I type, red) at the rt269 amino acid position. Ten sequences each of subgenotypes C2(1), C2(2), and C2(3) have been added, as well as HBV reverse transcriptase sequences from patients (B182, B184) with coexisting HBV rt269L type and rt269I type.

Geographical distribution of the rt269L and rt269I types of subgenotype C1

In subgenotype C1, the rt269L and rt269I types diverged early in their evolutionary history, suggesting a disparity in geographic origin between them (Fig. 4). To address this hypothesis, we examined the geographical distribution between rt269L and rt269I of subgenotype C1, which is distributed mainly in East Asian and Southeast Asian countries (Fig. 5). Out of 408 subgenotype C1 HBV sequences, 400 HBV sequences (196 rt269L and 202 rt269I) with available geographical information were used (**Supplementary Table 3**). Among South Asian nations, the rt269L type is predominant in the western region of Southeast Asia, including Thailand (54strains/63, 85.7%) and Bangladesh (30strains/32, 93.8%), and the rt269I type is more prevalent in East Asia, including Vietnam (41strains/41, 100%), China (24strains/36, 66.7%), and Hong Kong (43strains/46, 93.5%). In Cambodia [L: 47.2% (34strains /72) vs. I: 51.4% (37strains /72)] and Malaysia [L: 54.5% (30strains/55) vs. I: 45.5% (25strains/55)], the rt269L and rt269I types are almost similarly distributed. Together, our data indicate that the rt269L and rt269I types of subgenotype C1 have distinct geographical origins in Asia.

Distinct distribution of the basal core promoter and precore gene mutations in rt269L and rt269I of subgenotype C

In the 3215 bp HBV genotype C genome, we identified a total of 20 signature mutations. Among these sequences, 17 were synonymous mutations, and three were nonsynonymous (NS) mutations located in the spacer region of the polymerase, including Spacer S196P, Spacer V214I and Spacer A289T (Table 4). Among the 17 synonymous mutations, there was a distinct frequency difference of the T1858-C mutation in the precore region between the rt269L and rt269I types of subgenotype C1; this difference is related to resistance to the 1896 G to A precore mutation via the stem loop structure of pregenomic RNA and liver disease progression in genotype C[29]. Most cases of the rt269L type (186/197, 94.4%) encode thymidine at position 1858 (T-1858), whereas most cases of the rt269I type (137/209, 65.6%) encode cytosine (C-1858). However, no disparity in the frequency of the T1858C

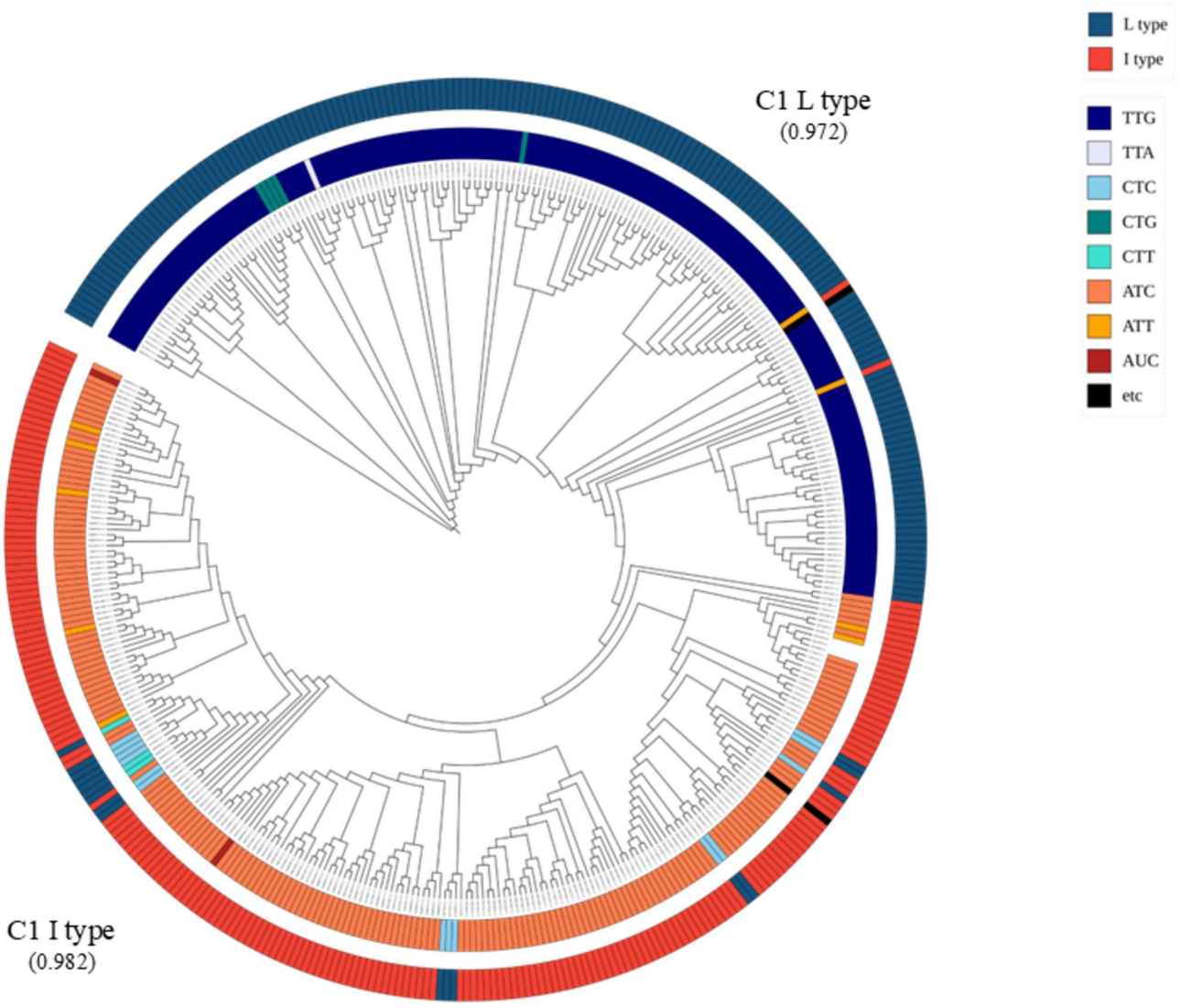

Fig. 4. Phylogenetic divergence of rt269L and rt269I in subgenotype C1. This phylogenetic tree was constructed with 408 HBV subgenotype C1 and 3 genotype A as outgroups. This figure shows that rt269L and rt269I in subgenotype C1 have diverged since the beginning of evolution. The inner ring represents nucleotide triplets, color-coded according to the legend on the right (e.g., TTG, TTA, CTC). The outer ring categorizes these sequences into two primary types: the L type, depicted in blue, and the I type, depicted in red. The support value is listed beneath the name of each main clade.

mutation between the two types was found in subgenotype C2, with T1858 being predominant in both the rt269L and rt269I types (Table 5).

Since the T1858C mutation can affect the 1896 precore mutation related to the inhibition of HBeAg production[30], we investigated the frequency of the 1896 precore mutation in the rt269L and rt269I types. We found that the frequency of the 1896 precore mutation was significantly greater in the rt269L type than in the rt269I type in subgenotype C1 [15.2% (30/197) vs. 8.6% (18/209)], whereas subgenotype C2 presented a greater frequency of the 1896 precore mutation in the rt269I type. The frequency of the BCP double mutations at 1762 and 1764 was significantly greater in the rt269I type in subgenotype C1 than in the rt269L type [41.1% (81/197) vs. 56.9% (119/209)] as in genotype C2 [57.8% (193/334) vs. 63.0% (218/346)]. Taken together, our data indicates that there is a distinct distribution difference in the basal core promoter (A1762T/G1764A) and precore (T1858C and G1896A) gene mutations between the rt269L and rt269I types of subgenotype C.

Discussion

In this study, we investigated the phylogenetic distribution, geographic separation, and mutation frequency of the rt269L and rt269I types in HBV subgenotype C1 and compared our findings with those in subgenotype C2. Several key insights emerged from our analysis.

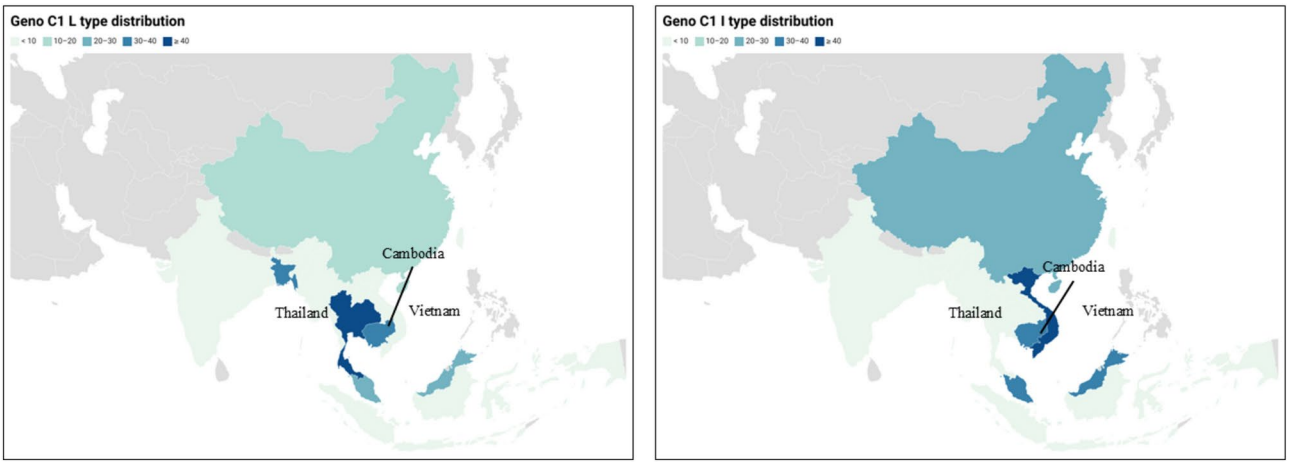

Fig. 5. Geographical distribution of the rt269L and rt269I types in HBV subgenotype C1. This figure illustrates the geographical distribution of leucine (L) and isoleucine (I) types at the rt269 site within the HBV subgenotype C1 across Asia. In the left panel, darker blue indicates higher frequencies of the L variant in Thailand, Cambodia, and Bangladesh. Similarly, the right panel uses darker blue to denote higher frequencies of the I variant in Vietnam, Hong Kong, and China.

Our phylogenetic analysis revealed a clear separation between the rt269L and rt269I types in subgenotype C1, unlike the mosaic evolution observed in subgenotype C2. This distinct phylogenetic clustering suggests that the rt269L and rt269I types in subgenotype C1 followed different evolutionary paths. The significant difference in codon sequences between rt269L (TTG) and rt269I (ATC) in subgenotype C1 further supports this separation. In contrast, within subgenotype C2, the mutation barrier was low, as rt269L(CTC) and rt269I(ATC) differ by only a single nucleotide. Phylogenetic analysis of HBV DNA extracted from patients infected with subgenotype C2, in which both rt269L and rt269I types were detected, indicates that these two variants do not represent distinct infections but rather quasispecies resulting from mutation.

The early phylogenetic divergence between the rt269L and rt269I variants implies the existence of a temporal or spatial separation between the two types. Geographically, the rt269L and rt269I types in subgenotype C1 have distinct distributions, with rt269L being predominant in Southwest Asia (e.g., Thailand and Bangladesh) and rt269I being more common in East Asia and Southeast Asia (e.g., Vietnam, China, and Hong Kong)[31–33]. This geographic segregation is consistent with the phylogenetic findings and suggests that regional factors, such as countries, may have influenced the evolution and distribution of these polymorphisms[34]. For instance, Cambodia, located between Vietnam, where rt269I is common, and Thailand, where rt269L is prevalent, displays both rt269L and rt269I in roughly equal proportions.

Our study identified 20 signature mutations that differentiate the rt269L and rt269I types in subgenotype C1, including significant differences in the frequencies of precore (T1858C/G1896A) and basal core promoter (A1762T/G1764A) mutations. The T1858C mutation, which is related to resistance to the 1896 precore mutation and greater liver disease severity[35], was more frequent in the rt269I type. This result is consistent with previous findings that link the C-1858 mutation to severe chronic liver disease, particularly in Vietnam, where all C1 patients studied had the rt269I type[36]. Conversely, the 1896 precore mutation, which inhibits HBeAg production, was more common in the rt269L type. The higher frequency of the T1858C mutation in the rt269I type may contribute to the lower prevalence of the 1896 precore mutation in these strains, as T1858C prevents the formation of the G1896A mutation through its impact on the pregenomic RNA encapsidation signal's stem–loop structure[29]. The higher frequency of basal core promoter (BCP) double mutations (A1762T/G1764A) in the rt269I type in both subgenotype C1 and subgenotype C2 suggests a propensity for more aggressive viral replication and disease progression compared to genotype A and B[37,38]. These mutations are known to increase viral replication and are associated with severe liver disease, further highlighting the clinical relevance of these findings[39,40].

The distinct phylogeographic and mutational characteristics of the rt269L and rt269I types in subgenotype C1 have important implications for HBV management and treatment strategies. Understanding the geographic and evolutionary dynamics of these polymorphisms can aid in the development of targeted therapies and vaccination strategies, particularly in regions where specific HBV subgenotypes and polymorphisms are prevalent[34,41]. Additionally, the differential mutation frequencies between the rt269L and rt269I types underscore the need for personalized approaches to HBV treatment[14,15].

This study has several limitations. First, the differences in disease characteristics or clinical outcomes between the rt269L and rt269I variants of subgenotype C1 are unknown, and therefore, it is unclear which clinical features may have driven the divergence between rt269L and rt269I types. Second, the relationship between rt269 variants and BCP and PC mutations could not be established, leaving the mechanisms by which rt269L and rt269I types acquire these mutations uncertain. Third, there were limitations related to sequence collection. For the other subgenotype C, the sample size was insufficient, and the collection dates for the sequences were unevenly distributed, which restricted the detection of a temporal signal.

Sub-genotype C1				L type (n = 198)				I type (n = 209)			
nucleotide position											
ORF (Pol)		nt1-nt3215		A	C	G	T	A	C	G	T
preS2	(Spacer)	Syn	70	172 (87.3%)	24 (12.2%)	0	0	3 (1.4%)	206 (98.6%)	0	0
SHB		Syn	796	0	0	30 (15.2%)	166 (84.3%)	5 (2.4%)	0	168 (80.4%)	35 (16.7%)
HBx	(RT)	Syn	1023	3 (1.5%)	178 (90.4%)	0	16 (8.1%)	0	18 (8.6%)	1 (0.5%)	190 (90.9%)
		Syn	1026	0	2 (1.0%)	0	195 (99.0%)	2 (1.0%)	146 (69.9%)	2 (1.0%)	59 (28.2%)
		Syn	1062	183 (92.9%)	0	14 (7.1%)	0	23 (11.0%)	0	186 (89.0%)	0
	(RNaseH)	Syn	1461	88 (44.7%)	41 (20.8%)	61 (31.0%)	3 (1.5%)	6 (2.9%)	25 (12.0%)	173 (82.8%)	2 (1.0%)
preC		Syn	1858	0	10 (5.1%)	0	186 (94.4%)	1 (0.5%)	137 (65.6%)	0	71 (34.0%)
Core		Syn	1975	4 (2.0%)	0	142 (72.1%)	50 (25.4%)	1 (0.5%)	0	6 (2.9%)	201 (96.2%)
		Syn	1984	51 (25.9%)	0	135 (68.5%)	10 (5.1%)	53 (25.4%)	1 (0.5%)	10 (4.8%)	144 (68.9%)
		Syn	2248	168 (85.3%)	3 (1.5%)	2 (1.0%)	24 (12.2%)	4 (1.9%)	0	4 (1.9%)	200 (95.7%)
	(TP)	Syn	2576	0	175 (88.8%)	0	21 (10.7%)	1 (0.5%)	7 (3.3%)	2 (1.0%)	199 (95.2%)
		Syn	2612	0	40 (20.3%)	0	157 (79.7%)	0	169 (80.9%)	0	40 (19.1%)
		Syn	2716	0	189 (95.9%)	0	8 (4.1%)	0	57 (27.3%)	0	149 (71.3%)
		Syn	2741	154 (78.2%)	0	43 (21.8%)	0	18 (8.6%)	0	191 (91.4%)	0
		Syn	2795	192 (97.5%)	1 (0.5%)	0	4 (2.0%)	61 (29.2%)	11 (5.3%)	0	136 (65.1%)
		Syn	2801	0	51 (25.9%)	0	146 (74.1%)	0	205 (98.1%)	0	4 (1.9%)
preS1	(Spacer)	(S196P)	2892	0	168 (85.3%)	0	27 (13.7%)	0	4 (1.9%)	0	205 (98.1%)
		(V214I)	2946	13 (6.6%)	0	181 (91.9%)	2 (1.0%)	194 (92.8%)	1 (0.5%)	13 (6.2%)	1 (0.5%)
		Syn	3065	50 (25.4%)	1 (0.5%)	145 (73.6%)	0	139 (66.5%)	0	69 (33.0%)	1 (0.5%)
		(A289T)	3111	4 (2.0%)	0	191 (97.0%)	2 (1.0%)	155 (74.2%)	0	51 (24.4%)	3 (1.4%)

Table 4. Signature mutations of the rt269L and rt269I types of subgenotype C1 and their frequencies. This table presents the frequency analysis of signature mutations according to the rt269 type in the HBV subgenotype C1. The nucleotide positions, type of mutation (synonymous (syn) or nonsynonymous), and frequency of each mutation in the L-type and I-type groups are listed. Parentheses indicate domains belonging to the HBV polymerase. RT stands for the reverse transcriptase of polymerase, and TP stands for the terminal protein of polymerase. For each nucleotide position, deletions and ambiguous sequences were excluded from the counts of nucleotides.

In conclusion, this study revealed significant phylogenetic, geographic, and mutational differences between the rt269L and rt269I types in HBV genotype C. Unlike C2, these types in C1 are distinct both phylogenetically and geographically, showing unique mutation frequencies, especially in the precore and basal core promoter regions. These differences highlight the need for further research and tailored HBV management strategies on the basis of subgenotype and polymorphism type.

			L type				I type				Chi-square P value
Genotype	position		A	C	G	T	A	C	G	T	
Geno C1	BCP mutation	A1762T	109 (55.3%)	1 (0.5%)	0	83 (42.1%)	81 (38.8%)	1 (0.5%)	1 (0.5%)	122 (58.4%)	0.0008
		G1764A	86 (43.7%)	0	104 (52.8%)	0	129 (61.7%)	0	77 (36.8%)	0	0.0005
		Double mutation	1762A_1764G		1762T_1764A		1762A_1764G		1762T_1764A		
		A1762T_G1764A	102 (51.8%)		81 (41.1%)		75 (35.9%)		119 (56.9%)		0.0009
	preC mutation	T1858C	0	10 (5.1%)	0	186 (94.4%)	1 (0.5%)	137 (65.6%)	0	71 (34.0%)	< 0.0001
		G1896A	30 (15.2%)	0	165 (83.8%)	0	18 (8.6%)	0	191 (91.4%)	0	0.0355
Geno C2	BCP mutation	A1762T	134 (40.1%)	0	0	196 (58.7%)	123 (35.5%)	0	0	221 (63.9%)	0.1950
		G1764A	199 (59.6%)	1 (0.3%)	130 (38.9%)	0	244 (70.5%)	0	101 (29.2%)	0	0.0051
		Double mutation	1762A_1764G		1762T_1764A		1762A_1764G		1762T_1764A		
		A1762T_G1764A	127 (38.0%)		193 (57.8%)		98 (28.3%)		218 (63.0%)		0.0222
	preC mutation	T1858C	0	11 (3.3%)	0	323 (96.7%)	0	8 (2.3%)	1 (0.3%)	337 (97.4%)	0.4414
		G1896A	85 (25.4%)	0	247 (74.0%)	0	123 (35.5%)	0	223 (64.5%)	0	0.0050

Table 5. BCP and precore mutation frequencies in the rt269L and rt269I types of subgenotypes C1 and C2. This table presents the distributions of the L and I types of BCP and precore mutations within each C1 and C2 genotype, along with the corresponding chi-square P values. For each nucleotide position, deletions or ambiguous sequences were excluded from the nucleotide count.

Data availability

The rt269 typing results for the patients, along with the accession numbers and metadata for all sequences used in this study, are presented in the Supplementary Material. Additionally, the assembled sequences have been deposited in NCBI GenBank. (accession nos. PQ008220 to PQ008234).

Received: 7 February 2025; Accepted: 12 March 2025
Published online: 18 March 2025

References

1. Schweitzer, A., Horn, J., Mikolajczyk, R. T., Krause, G. & Ott, J. J. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet* **386**(10003), 1546–1555 (2015).
2. Guvenir, M. & Arikan, A. Hepatitis B virus: From diagnosis to treatment. *Pol. J. Microbiol.* **69**(4), 391–399 (2020).
3. Tan, A. T., Koh, S., Goh, V. & Bertoletti, A. Understanding the immunopathogenesis of chronic hepatitis B virus: An Asian prospective. *J. Gastroenterol. Hepatol.* **23**(6), 833–843 (2008).
4. Liang, T. J. Hepatitis B: The virus and disease. *Hepatology.* **49**(5 Suppl), S13-21 (2009).
5. Nowak, M. A. et al. Viral dynamics in hepatitis B virus infection. *Proc. Natl. Acad. Sci.* **93**(9), 4398–4402 (1996).
6. Rajoriya, N., Combet, C., Zoulim, F. & Janssen, H. L. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B Time for an individualised approach?. *J. hepatol.* **67**(6), 1281–1297 (2017).
7. Kramvis, A. Genotypes and genetic variability of hepatitis B virus. *Intervirology* **57**(3–4), 141–150 (2014).
8. Kyaw, Y. Y. et al. Distribution of hepatitis B virus genotypes in the general population of Myanmar via nationwide study. *BMC Infect. Dis.* **20**(1), 552 (2020).
9. Velkov, S., Ott, J. J., Protzer, U. & Michler, T. The global hepatitis B virus genotype distribution approximated from available genotyping data. *Genes (Basel)* **9**(10), (2018).
10. Chan, H. L. et al. Hepatitis B virus genotype C takes a more aggressive disease course than hepatitis B virus genotype B in hepatitis B e antigen-positive patients. *J. Clin. Microbiol.* **41**(3), 1277–1279 (2003).
11. Gao, Z. Y. et al. Mutations in preS genes of genotype C hepatitis B virus in patients with chronic hepatitis B and hepatocellular carcinoma. *J. Gastroenterol.* **42**(9), 761–768 (2007).
12. Huy, T.T.-T. et al. Genotype C of hepatitis B virus can be classified into at least two subgroups. *J. gen. virol.* **85**(2), 283–292 (2004).
13. Chan, H. L. et al. Epidemiological and virological characteristics of 2 subgroups of hepatitis B virus genotype C. *J. Infect. Dis.* **191**(12), 2022–2032 (2005).
14. Tanaka, Y. et al. Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma. *J. Hepatol.* **45**(5), 646–653 (2006).
15. Lee, S.-Y. et al. rt269I type of hepatitis B virus (HBV) leads to HBV e antigen negative infections and liver disease progression via mitochondrial stress mediated type I interferon production in chronic patients with genotype C infections. *Front. Immunol.* **10**, 1735 (2019).
16. Jeong, H. et al. Rt269i type of hepatitis b virus (Hbv) polymerase versus rt269l is more prone to mutations within hbv genome in chronic patients infected with genotype c2: Evidence from analysis of full hbv genotype c2 genome. *Microorganisms.* **9**(3), 601 (2021).
17. Ahn, S. H. et al. Substitution at rt269 in hepatitis B virus polymerase is a compensatory mutation associated with multi-drug resistance. *PloS one.* **10**(8), e0136728 (2015).

18. Park, E.-S. et al. Identification of a quadruple mutation that confers tenofovir resistance in chronic hepatitis B patients. *J. Hepatol.* **70**(6), 1093–1102 (2019).
19. Choi, Y.-M., Kim, D. H., Jang, J., Choe, W. H. & Kim, B.-J. rt269L-Type hepatitis B virus (HBV) in genotype C infection leads to improved mitochondrial dynamics via the PERK–eIF2α–ATF4 axis in an HBx protein-dependent manner. *Cell. Mol. Biol. Lett.* **28**(1), 26 (2023).
20. Kim, K. et al. Locked nucleic acid real-time polymerase chain reaction method identifying two polymorphisms of hepatitis B virus genotype C2 infections, rt269L and rt269I. *World J. Gastroenterol.* **29**(11), 1721–1734 (2023).
21. Kim, J. E. et al. Naturally occurring mutations in the reverse transcriptase region of hepatitis B virus polymerase from treatment-naïve Korean patients infected with genotype C2. *World J. Gastroenterol.* **23**(23), 4222–4232 (2017).
22. McNaughton, A. L., Revill, P. A., Littlejohn, M., Matthews, P. C. & Ansari, M. A. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. *J. Gen. Virol.* **101**(3), 271–283 (2020).
23. Kim, D. H., Choi, Y. M., Jang, J. & Kim, B. J. Global prevalence and molecular characteristics of three clades within hepatitis B virus subgenotype C2: Predominance of the C2(3) clade in South Korea. *Front. Microbiol.* **14**, 1137084 (2023).
24. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evolut.* **30**(4), 772–780 (2013).
25. Mulyanto, et al. A nationwide molecular epidemiological study on hepatitis B virus in Indonesia: Identification of two novel subgenotypes, B8 and C7. *Arch. virol.* **154**, 1047–1059 (2009).
26. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Boil. evol.* **26**(7), 1641–1650 (2009).
27. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucl. Acid. res.* **49**(W1), W293–W296 (2021).
28. Kumar, S., Tamura, K. & Nei, M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics.* **10**(2), 189–191 (1994).
29. Lindh, M., Furuta, Y., Vahlne, A., Norkrans, G. & Horal, P. Emergence of precore TAG mutation during hepatitis B e seroconversion and its dependence on pregenomic base pairing between nucleotides 1858 and 1896. *J. Infect. Dis.* **172**(5), 1343–1347 (1995).
30. Park, Y. M. & Lee, S. G. Clinical features of HBsAg seroclearance in hepatitis B virus carriers in South Korea: A retrospective longitudinal study. *World J. Gastroenterol.* **22**(44), 9836–9843 (2016).
31. Sugauchi, F. et al. Hepatitis B virus genotypes and clinical manifestation among hepatitis B carriers in Thailand. *J. gastroenterol. hepatol.* **17**(6), 671–676 (2002).
32. Chan, H. L. Y. Significance of hepatitis B virus genotypes and mutations in the development of hepatocellular carcinoma in Asia. *J. gastroenterol. hepatol.* **26**(1), 8–12 (2011).
33. Truong, B. X. et al. Variations in the core promoter/pre-core region in HBV genotype C in Japanese and Northern Vietnamese patients. *J. Med. Virol.* **79**(9), 1293–1304 (2007).
34. Holmes, E. C. Evolutionary history and phylogeography of human viruses. *Annu. Rev. Microbiol.* **62**(1), 307–328 (2008).
35. Lindh, M., Andersson, A.-S. & Gusdal, A. Genotypes, nt 1858 variants, and geographic origin of hepatitis B virus—large-scale analysis using a new genotyping method. *J. Infect. dis.* **175**(6), 1285–1293 (1997).
36. Huy, T. T. T. et al. Characteristics of core promoter and precore stop codon mutants of hepatitis B virus in Vietnam. *J. med. virol.* **74**(2), 228–236 (2004).
37. Zhang, D. et al. Prevalent HBV point mutations and mutation combinations at BCP/preC region and their association with liver disease progression. *BMC infect. Dis.* **10**, 1–8 (2010).
38. Zhong, Y. et al. Hepatitis B virus basal core promoter/precore mutants and association with liver cirrhosis in children with chronic hepatitis B virus infection. *Clin. Microbiol. Infect.* **22**(4), 379.e1-e8 (2016).
39. Kumar, R. Review on hepatitis B virus precore/core promoter mutations and their correlation with genotypes and liver disease severity. *World J. Hepatol.* **14**(4), 708 (2022).
40. Yotsuyanagi, H. et al. Precore and core promoter mutations, hepatitis B virus DNA levels and progressive liver injury in chronic hepatitis B. *J. Hepatol.* **37**(3), 355–363 (2002).
41. Lin, Y.-Y. et al. New insights into the evolutionary rate of hepatitis B virus at different biological scales. *J. Virol.* **89**(7), 3512–3522 (2015).

Author contributions

B-JK and DK designed the study. B-JK interpreted the research and wrote the first draft of the manuscript. DK performed the data analysis and revised the manuscript. Y-MC and JJ supported the data analysis and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Research Foundation of Korea (NRF), funded by the Ministry of Education (Grant No. RS-2025–00553721). Yu-Min Choi received funding from the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (Grant No. RS-2023–00276686).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94286-6.

Correspondence and requests for materials should be addressed to B.-J.K.