

Evolutionary analysis of pre S/S mutations in HBeAg-negative chronic hepatitis B with HBsAg <100 IU/ml

Yingping Wu (✉ yingpingwu@zju.edu.cn)

Zhejiang University <https://orcid.org/0000-0003-3666-1313>

Zhiqiang Zhu

Zhejiang University School of Medicine

Jianyong Wu

Zhejiang University School of Medicine

Wenzi Bi

Zhejiang University School of Medicine

Weigen Wu

Zhejiang University School of Medicine

Wei Xu

Zhejiang University School of Medicine

Research

Keywords: Inactive HBsAg carrier, Reactivation phase, Genome diversity, Pre-S/S region mutation

Posted Date: April 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20997/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Hepatitis B surface antigen (HBsAg) and viral load are important clinical indicators for patients with chronic hepatitis B (CHB) infection treated with antiviral therapy. Few studies have evaluated viral sequence biomarkers predicting the risk of hepatocellular carcinoma (HCC) in the stage, which show a low serological response (HBsAg <100 IU/ml) and high viral levels (HBV DNA >2,000 IU/ml). Additionally, point mutations within the pre S/S regions of HBsAg have frequently been reported to be associated with diagnostic failure, vaccine escape and immunotherapy escape. However, the prevalence of escape mutations with low levels of HBsAg (<100 IU/ml) in inactive HBsAg carriers has not been systematically studied within the past decade.

Methods: This study aims to determine the trend of the biological prevalence of escape mutations within the pre S/S regions of special model of inactive CHB infection. Increased diversity over the complete HBV genome at baseline was associated with evolutionary characteristics. We used Sanger sequencing, quantitative HBV serology (HBeAg and HBsAg) and *liver function index* (aspartate alanine aminotransferase, ALT; alanine aminotransferase, AST) to identify whether HBV genome sequences are associated with the serological response to long-term host immunity in special inactive CHB infection.

Results: Compared to HBV in HCC, HBV sequencing analysis of 28 LR/RA CHB patients with genotypes B and C showed higher genetic diversity among four ORFs, but the two groups presented similar selective pressure. Seven positive selection sites in the pre-S1 region are potentially associated with immune evasion. These mutations in the pre-S/S region might be associated with the HCC phenotype of low HBsAg expression, with the P region possibly impacting high viral loads.

Conclusion: According to the results of this study, LR/RA CHB is characterized by not only genetic diversity but also positive evolutionary pressure within the pre-S/S regions. Increased viral diversity across the HBV genome is also associated with low levels of HBsAg. In conclusion, the cumulative evolutionary changes in the HBV pre-S/S regions that facilitate immune evasion should be monitored individually. Due to the similarity of evolutionary characteristics in HCC, low serological responses and high viremia may be associated with the risk of HCC.

Introduction

Hepatitis B virus (HBV) infection has a wide spectrum of clinical manifestations ranging from an asymptomatic carrier state (immunotolerant state) to acute or chronic hepatitis, with progression to severe liver disease [1]. Although available therapies are highly effective at controlling viral replication, they often involve life-long treatment because infections with HBsAg seroconversion are very rarely cured.

Previous studies have reported that the serum HBsAg level is related to intrahepatic covalently closed circular DNA (cccDNA) [2]. Lowering of serum HBsAg to an undetectable level may indicate that intrahepatic cccDNA has been eradicated, which is known as a “functional cure”; however, this state is difficult to achieve through current antiviral approaches [3, 4]. At the same time, serum HBV DNA levels

may indirectly reflect immunological control of HBV infection independent of the HBV DNA viral load, which is considered to represent viral replication activity [5]. In the present study, we focused on a special infectious pattern between the low replicative phase (LR, also referred to as the “inactive HBsAg carrier” state) and the reactivation phase (RA, previously also referred to as “HBeAg-negative/anti-HBe positive chronic hepatitis B”) CHB that is characterized by a low serological response (HBsAg < 100 IU/ml) and high viral levels (HBV DNA > 2,000 IU/ml), persistently normal aminotransferase levels and mild inflammation and minimal fibrosis in the liver; this pattern is herein referred to as LR/RA CHB.

Recent studies have demonstrated that serum HBsAg levels are a highly predictive factor of a sustained outcomes due to the nature and strength of the host immune response against HBV [6]. Monitoring qHBsAg levels may aid in tracking the natural history of the disease and in predicting the response to antiviral treatment and natural immune clearance [7]. Due to low HBsAg levels, the incidence of HBsAg loss in high genotypes B, C and D patients in CHB is linked with a lower risk of HCC [8]. In this study, we evaluated whether qHBsAg levels can predict a lower risk for HCC in patients with HBV DNA > 2000 IU/ml and HBsAg < 100 IU/ml.

The diversity of the HBV genome is important for individualized therapies for CHB patients. HBsAg has a significant impact on the performance of diagnostic screening tests and the clinical outcome of hepatitis B infection. Previously, we identified various mutations in patients with a high serum viral load by utilizing sequencing and bioinformatic analysis. Compared with HCC, the HBV genome, including the P/X/S/C regions, exhibits greater diversity, and positive selective pressure is observed within the pre-S/S region. These mutations have been shown to occur within the “a” determinant region of the S region associated with HCC. Nevertheless, the interaction between pre-S/S region mutations and host immunity in LR/RA CHB is still unknown. The aim of the present study was to examine whether this infectious module from LR to RA is related to the risk of developing HCC. We clarify the association between the evolutionary diversity of the HBV genome (especially pre-S/S mutations) and disease stage under LR/RA CHB, screening for new molecular markers of HCC. Furthermore, this study provides novel insight into the mechanism of HBV-related HCC.

Materials And Methods

Patients. For this study, 28 patients with chronic HBV infection were enrolled from the Fourth Affiliated Hospital of Zhejiang University of Medicine in Yiwu. The patients exhibited RA CHB and were treatment-naïve. All patients met the following criteria: positive for the HBV surface antigen (< 100 IU/ml), HBV DNA levels > 10³ IU/ml, ALT/AST levels (< 50 U/liter), and negative for HBeAg, for antibodies indicating human immunodeficiency virus, hepatitis C virus, or hepatitis virus coinfection and for decompensated liver disease. No patients received nucleotide/nucleoside analog therapy prior to blood sampling. Written informed consent was obtained from all patients, and the study protocol was approved by the Ethics Committee of Zhejiang University. The experimental methods were carried out in accordance with approved guidelines.

Liver biochemistry, HBV serology, and HBV DNA tests. Liver biochemical (ALT, AST and total bilirubin) parameters were tested using an automated chemistry analysis system (Beckman Coulter, CA, USA). HBV serological markers were determined with a chemiluminescent microparticle immunoassay using the Abbott Architect immunoassay system (Abbott Laboratories, IL, USA). HBV DNA levels were measured by PCR using a Cobas z480 system (Roche Diagnostics, Mannheim, Germany), with a low limit of quantification of 100 IU/ml. Only patients with RA CHB with serum HBV DNA levels greater than 10^3 IU/ml and HBsAg levels lower than 100 IU/ml were included in the study.

Extraction of viral DNA and genome-length PCR. HBV DNA was extracted from 200 μ l of serum samples from patients with RA CHB using QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. We performed seminested PCR to amplify four partially overlapping DNA fragments that encompass the complete HBV genome sequence using HotStarTaq Plus DNA Polymerase (QIAGEN, Hilden, Germany), primers, and conditions as previously described. The PCR products were then sequenced bidirectionally with second-round primers using an Applied Biosystems 3730XL (Applied Biosystems, CA, USA). The full-length HBV genomic DNA sequence was amplified by an Applied Biosystems 2720 thermal cycler (Applied Biosystems, CA, USA) in a 25 μ l volume containing 1 μ l of HBV DNA template, 0.5 μ l of Taq DNA polymerase (5U/ μ l), 2.5 μ l of 10x Taq buffer, 2 μ l of 25 mM MgCl₂, 0.5 μ l of dNTPs (10 mM each), and 1 μ l of each primer. The primers were designed according to the corresponding reference [27]. The total genomes of 28 viral strains from the 28 patients were sequenced.

Sequence analysis. DNASTAR software (DNASTAR Inc., Madison, WI, USA) was used to assemble the sequences into a complete HBV genome sequence, and all mutations were refined manually. HBV genotyping and phylogenetic analysis were performed using Mega 7.0.

Phylogenetic analysis. Phylogenetic trees were reconstructed from either the full-length HBV genomes or concatenated nucleotide alignments (four protein-coding: Pol, LHBsAg, HBxAg, and PreC/C). We employed ModelTest 3.7 to select the (GTR + G) model of sequence evolution for ML analysis, under the Akaike information criterion. PhyML 3.0 was then applied to create maximum likelihood phylogenies with 1000 bootstrap replicates. MrBayes 3.1.2 was employed for Bayesian inference, which was conducted twice with four Markov chains each time. The runs were performed for 5×10^6 generations, with sampling every 100 generations. When the log likelihood scores were found to stabilize, a consensus tree was calculated after omitting the first 25% of the trees as burn-in. As the two phylogenetic methods generated similar topologies, only the MrBayes phylogeny is presented.

Sequence diversity. Sequence diversity was calculated using Shannon entropy (Sn), which measures the diversity of the number of haplotypes and their frequencies. The diversity of each nucleotide position (nt1 to 3215) was calculated as the Shannon entropy [$Sn = -\sum_i [A, T, C, G, -] (p_i \ln p_i) / \ln N$], where pi represents the relative frequency of nucleotides or deletion at this position and N the total number of sequences [28]. The figure illustrating diversity (Fig. 2) was evaluated in MEGA 7.0 [29] with three parameters, the mean genetic distance (d), the number of synonymous substitutions per synonymous site (dS), and the number of non-synonymous substitutions per nonsynonymous site (dN); it was generated using the Circos

visualization tool [30]. All samples were compared with the HCC reference indicated above to identify variations in ORFs at the amino acid level.

Analysis of selection pressure and positive selection. We used the dN/dS ratio, a measure of selection, to assess the selection pressure acting on a lineage of HBV. Dn, dS and the dN/dS ratio were estimated using the codeML model in PAML4 [31]. Lineage-specific mean values were estimated with concatenated alignments of all orthologs. The KaKs calculator was used to calculate dN/dS in a sliding window and scale (<http://www.bacteriamuseum.org/SWAAP/SwaapPage.htm>) using the Nei- Gojobori distance estimation method (window length, 57 nt; window step, 6 nt). The direction of selection pressure was determined by measuring the variable ω , representing nonsynonymous/synonymous substitution ratios ($\omega = dN/dS$) at each codon site, with values of $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicating purifying selection, neutral evolution, and positive selection, respectively. The extent of positive selection was analyzed using a site model employing two different pairs of models (M1/M2, M7/M8). Model M1 assumes negative selection and neutral evolution; model M2 assumes an additional level of positive selection. The M7 and M8 model pair assumes beta distributions for ω among sites, providing a sensitive test for positive selection. Likelihood ratio tests were utilized to compare the nested models, and empirical Bayes methodology was applied to identify the amino acid sites under positive selection due to a more reliable posterior probability calculation for small datasets [32]. Finally, PyMOL 2.3 software was used to show amino acid variants in HBV epitopes with a three-dimensional conformation.

Statistical analysis. The distribution of point mutations within the pre-S/S regions at amino acid substitutions within the large HBsAg antigen in the experimental group was evaluated with a chi-square test using SPSS 19.0. Sequence complexity or diversity was analyzed with Student's t-test, and p-values of < 0.05 were considered significant.

Results

Cohort characteristics. HBeAg-negative CHB patients with coexisting low HBsAg levels (< 100 IU/ml) and high HBV DNA levels ($> 2,000$ IU/ml) ($n = 28$) exhibited similar characteristics and metabolite profiles, including their mean age, sex, alanine aminotransferase (ALT) levels, aspartate transaminase (AST) levels, hepatitis B surface antigen (HBsAg), and hepatitis B DNA (HBV DNA) (Table 1). The experimental group was in the reactivation phase. HBV-related HCC sequences from NCBI were present in the control group. The HBV DNA level in the experimental group was significantly higher than that in the regular reactivation phase ($> 2,000$ IU/ml), but with low levels of HBsAg (< 100 IU/ml) and ALT, which are not common. In total, 28 full-length HBV genomes were analyzed for viral diversity, phylogenetic divergence, selection pressure, and positive selection.

Phylogenetic analysis. We were interested in identifying molecular evolutionary characteristics specific to LR/RA CHB, and employed population sequencing to analyze the relationship between LR CHB and HCC. The reference sequences of the entire HBV whole genome from HCC patients for HBV genotyping were downloaded from NCBI as the control group; the accession numbers are EU522069, AF182805,

AB014368, EF137802, AB014360, AB014367, AB670237, AY206374, JQ429081, JQ429080, JQ027328, JQ027334, EU660230, AY206380, AB014366, AY206377, EU919174, EU522073, AY206375, AY206373, EU919161, JQ027329, EU881997, AY206383, JQ027315, EU564822, EU882003, EU919175, AY206391, EU882001, EU919170, EU919172, EU487257, EU564825, EU564825, JQ027331, JQ027330, EU522067, AY206390, JQ027325, AY206387 and EU487256. Due to different prognosis risks among HBV genotypes, the HCC control group was divided into genotypes B and C; the experimental group was similarly divided. Based on construction of phylogenetic tree, the analytical data indicated the presence of HBV genotypes B and C in the experimental group (Fig. 1). In general, all sequences clustered together according to genotype, and there were 22 strains (75%) of genotype B and 6 strains (25%) of genotype C. No other genotypes were detected. The identified HBV genotypes of group I (A6, A7, A10, A12, A17, A26, HBV DNA, > 2000 IU/ml and < 20000 IU/ml) and group II (B1, B3, B4, B5, B6, B7, B8, B9, B12, B13, B15, B16, B17, B19, B21, B23, B24, B25, B26, B27, B28, B29, B30, HBV DNA, > 20000 IU/ml) are the main common genotypes in China. The results of subgenotyping analysis are illustrated in Fig. 1(B), and the reference sequences of HBV subgenotypes are AB010289, AP011084, AB115551, AB287314, DQ463787, AB031265, AF473543, AF182804, AF182804, AB644284 and AB031262. From the identity matrix, genotype B of our experimental group belongs to HBV B2 (GenBank accession number AP011084); genotype C of our experimental group shares high identity with HBV C2 (GenBank accession number AY278488.2).)

Diversity of HBV sequences in RA CHB. Comparable sequence diversity was found among the full-length HBV genomes in RA CHB (Fig. 2) and 4 ORFs (opening reading frames) of HBV genotypes B and C (Fig. 3). The Shannon entropy (S_n) values of nucleotides and amino acids for the full-length HBV genome (only at the nucleotide level) and the P/C/S/X genes encoding LHBsAg, MHBsAg, HBsAg, HBxAg, HBcAg, Pol, reverse transcriptase (RT), and the core promoter (CP) region (only at the nucleotide level) were significantly greater in the experimental group than in the control group. The sequence distribution diversity of the P/C/S/X regions was significantly different between the experimental and control groups according to Shannon entropy analysis by R code (Fig. 3). Compared to the control group, the experimental group exhibited significantly higher Shannon entropy values ($p < 0.05$) at both the nucleotide and amino acid levels. Different from the control group of HCC patients, the sequences diversity of the P and X regions was significantly different in the experimental group, regardless of genotype B or C. The control group typically exhibited fewer mutations than did the experimental group, with an exceptional number of mutations found in the genes encoding LHBsAg, MHBsAg, and HBsAg in genotypes B and C, located in HLA I T cell epitopes or multiple types of epitopes (Fig. 2).

Selective pressure. The dN/dS ratio was used to measure selective pressure across individual lineages established with 28 experimental group sequences and control group (HCC group) sequences. The nonsynonymous/synonymous rate ratio ($\omega = dN/dS$), shows selective pressure at the protein level. If $dN < dS$ (then $\omega < 1$) is calculated, purifying selection will reduce their fixation rate. If $dN > dS$ is favored by Darwinian selection, they will be fixed at a higher rate so that $\omega > 1$ and is considered as adaptive protein evolution. We utilized the free ratio model (M1) in PAML4 to calculate the dN/dS ratio for 4 protein-coding genes (Pol, LHBsAg, HBxAg and PreC/C) individually, and the variability of dN/dS in the pre-S/S region was further detailed using sliding window analysis (Fig. 4A, 4B). Overall, the 4 protein-coding genes in the

two study groups exhibited negative selection and purifying selection reduced their fixation rate such that dN / dS < 1. However, the dN/dS ratio of the P/X regions was significantly higher in the control group (HCC) than in the experimental group (0.1802 ± 0.0003 , $p < 0.001$; 0.0827 ± 0.0011 , $p < 0.001$). In contrast, the dN/dS values of the pre-S/S region showed little difference between the experimental group and control group (0.1859 ± 0.0006 vs 0.1534 ± 0.0004 , $p = 0.06$). The sequence mutations between the P and X regions are related to HCC progression, and the study group may have a low risk. Moreover, based on dN/dS values in pre-S/S region, more positive selection between nucleotide positions 313 and 547 occurred in the control group, whereas significant selection pressure between nucleotide positions 859 and 1015 occurred in the experimental group. Therefore, it seems that the sequences of experimental group will evolve differently from those of the control group, avoiding the risk of further HCC progression.

Positive selection. A codon-based molecular evolution model was used to identify positive selection in the HBsAg protein-coding gene. In both groups, HBsAg was under positive selection according to the M8 model. In the experimental group, LHBsAg, MHBsAg, and HBsAg were all under positive selection according to the M8 model. Seven positively selected sites were identified in pre-S1, 5 in pre-S2 and 4 in S using the empirical Bayes method (Table 2). These sites of positive selection among pre-S/S regions improved the adaptability of HBV and were conducive to its survival.

Homologous modeling analysis of positive selection in HBsAg. The amino acid sequence variations of HBsAg were modeled by using PyMOL software (Fig. 5). The six sites of positive selection found in HBsAg are mainly located in antigen epitopes.

Discussion

The coexistence of lower levels of HBsAg and higher levels of HBV DNA was a unique serological profile identified in this study. Low replication activity was observed in the tumors of HCC patients, due to the low HBV pgRNA-to-DNA ratio, similar to the control group (HCC group). Previous studies have shown that the levels of HBsAg correlate with HBV DNA and intrahepatic covalently closed circular DNA (cccDNA) levels. However, some studies have found that serum HBsAg levels correlate weakly or moderately with serum HBV DNA and are weakly or not associated with intrahepatic cccDNA. Low HBsAg can predict subsequent HBsAg loss and the risk of HCC [9]. The combination of HBsAg and HBV DNA suppression is now considered as the most important endpoint for clinical trials and the ultimate short-term goal of treatments with the medications currently available [10]. Thus, chronic HBsAg carriers may simultaneously experience a lowering of the viral load to < 2000 IU/ml and the HBsAg level to < 100 IU/ml, reducing the risk of liver cancer [11]. The purpose of our study was to investigate whether CHB patients in the LR/RA phase with special serological profiles require monitoring management.

The natural history of HBV infection can generally be divided into four periods: the immune tolerance phase, immune clearance phase, inactive phase and reactivation phase. The demographic and clinical features of our study subjects were in accordance with the classification criteria for the reactivation phase (Table 1). Previous studies have suggested that patients with inactive HBV infections have a lower

risk of developing liver cancer because they exhibit lower levels of HBsAg (< 1000 IU/mL) [12] and there is a probability of spontaneous clearance. In addition, a quantitative HBsAg level less than 100 IU/ml can be used as the optimal threshold for predictive clearance [13], but no consensus regarding the exact value of the reference endpoint as an alternative threshold in clinical treatment has been reached [14]. Some studies have shown that most patients with low-level viremia (LLV; <2000 IU/mL) in the inactive phase of CHB infection present minimal evidence of liver injury [15]. However, we focused on a CHB cohort with high-level viremia but low-level HBsAg in this study (HBV DNA > 20000 IU/ml, 42.5%). An undetectable viral level rather than loss of HBsAg is associated with a lower risk of HCC [16]. Determination of whether the elimination of HBsAg as the endpoint of long-term antiviral treatment reduces, or merely delays the occurrence of liver cancer will require the discovery of more biomarkers to support the explanation. A unique feature of LR/RA CHB observed in this study was that, unlike for the wild-type strain, the serum HBsAg titer was not related to the HBV DNA load, and there was a possibility of immune escape. Consistent with previous studies, our results showed a higher level of full-length HBV genomic diversity in the experimental group than in HCC patients.

Moreover, HBV whole-genome sequences showed greater diversity in LR/RA CHB with lower levels of HBsAg and higher levels of HBV DNA than in the HCC control group. Additionally, the mean Shannon entropy score was not different between genotype C and genotype B subjects. Although we did not identify any hot-spot mutations or pre-S deletions associated with HCC, we found a higher frequency of P/S/X/pre-C/C mutations in LR/RA CHB based on the mean Shannon entropy. Previous studies have shown that mutation in the pre-S/S region affects HBsAg expression, with the final manifestation being the immune response between the virus and host; that is, the stronger is the immune response, the lower is the HBsAg level [17]. The low serum level of HBsAg observed in HBsAg-negative CHB (< 100 IU/mL) suggests a high probability of negative HBsAg conversion [18]; on the other hand, a negative level of HBsAg serology suggests that the virus integrates into the host genome to some degree. The analysis of 28 full-length genome sequences from 28 RA CHB patients infected with genotypes B and C (Fig. 1) showed that the viral diversity of protein-coding genes (LHBsAg, HBxAg, PreC/C, and Pol) was significantly greater in the experimental group than in the HCC group (Fig. 2). This finding was consistent with the higher nucleotide Sn (frequency of mutated positions) and Dn values for these genes in the experimental group ($p < 0.001$) (Fig. 3). Previous studies have suggested that increased diversity over the entire HBV genome at baseline is associated with reduced HBsAg loss under NA therapy in patients with genotype A and D, but our experimental group consisted of genotype B and C patients with LR/RA CHB who did not receive NA therapy. Regardless, the mechanism underlying the impact of whole-genome diversity on low HBsAg and high viral replication in LR/RA CHB remains unclear.

The dN/dS ratio is used to measure the strength of selection acting on protein-coding genes. To estimate the immune selection pressure on the HBV genome, we calculated the dN/dS ratios of protein-coding genes (LHBsAg, HBxAg, PreC/C, and Pol). The dN/dS ratios of two genes (HBxAg and polymerase) were significantly lower in the experimental group than in the control group ($p < 0.001$), indicating relatively strong negative selection in the former compared to the latter. However, the dN/dS ratios of the other two genes (HBsAg and PreC/C) in the experimental group were not significantly different from those in the

HCC group. In CHB, the frequency of mutations (or deletions) in the pre-S/S region caused by natural selection or antiviral therapy is very high [19], and the mutation (or deletion) pattern is not only detected in cirrhotic tissues. More commonly, in liver cancer tissue samples [20], pre-S mutations increase the incidence of HCC by 377-fold [21], which has been confirmed in several prospective cohort studies [22]. In addition, S region mutations are associated with liver fibrosis and liver cancer due to HBV RNA splicing [23]. Similar to the control group, our results suggest immune escape of the S gene in the experimental group, but the diversity of X gene mutations related to HCC was the same. Subsequently, the dN/dS of the pre-S/S region was verified by sliding window analysis using the KaKs calculator, and the results were in agreement with a previous study reporting that HCC patients have different regions of positive selection due to the special mode of CHB. This may lower the HCC risk in the experimental group and needs further study in the future. The results of our study were also in accord with the conclusion that CHB patients who ultimately achieve “functional cure” with loss of HBsAg, an uncommon occurrence with or without antiviral therapy, remain at risk for HCC development.

Due to the special phenotype of the patients in this study, we analyzed the codons of the HBV pre-S/S region that were under positive selection. Seven codons under positive selection were found in pre-S1, 5 in pre-S2 and 4 in S, which suggests that accumulating mutations may provide an opportunity for the virus to escape from host immune pressure. The expression level of HBsAg in RA CHB is usually low, but there are many factors influencing this phenotype, including nonspecific laboratory detection caused by viral gene mutations. In addition, the “a” determinant (amino acid residues 124–147) of the S region is an epitope recognized by the antibody, and point mutations in this determinant will result in a low affinity of the corresponding antibody [24]. In our study, 5 sites of positive selection with statistical significance were found in the “a” determinant (128H/135Q/135R/139L/141P), showing evolutionary diversity of HBV in the experimental group, which would be conducive to the survival of the virus. The combination of multiple mutations in the HBV pre-S/S region can also affect viral immunogenicity, along with the methodological limitations of existing commercial HBsAg quantification reagents [25], often delaying or interfering with clinical diagnosis and treatment. We further speculate that the pre-S/S mutant strains may lead to disproportionate synthesis and secretion of the virus such that the serum HBsAg titer does not reflect the viral replication ability [26]. The preliminary data from this study also indicated that the HBV genome sequence in LR/RA CHB exhibits more sites of positive selection in the pre-S/S region, and we must further study the interaction mechanism between the virus and host. As pre-S/S region mutations have been confirmed to be associated with HCC, we still need to evaluate the risk of pre S/S mutant strains for the progression of LR/RA CHB.

It is worth mentioning that the present study is the first to describe the characteristics of the full-length genome of the special mode with coexisting low levels of HBsAg and high levels of HBV DNA. However, our study has a few limitations. First, the case number was limited, and further large-scale studies are needed to confirm the results. Second, as this study used control sequences from databases, clinical information was not obtained. For example, age is an important factor to consider with regard to the immune response to HBV. Hopefully, the sites of positive selection identified in the pre-S/S region

showing significant differences can be used as candidate molecular markers, though further mechanistic studies are still required.

Conclusion

Understanding the evolution of pre-S/S mutations in patients with LR/RA CHB with a high viral load can increase our understanding of the pathogenesis of CHB. In our study, we showed that HBV viral diversity was increased in the experimental group, which was associated with increases in mutations in the pre-S/S region and the dN/dS of pre-S, followed by increases in HBV DNA levels and positively selected amino acids. These findings suggest that patients with coexisting low levels of HBsAg (< 100 IU/ml) and high levels of HBV DNA (> 2,000 IU/ml) may harbor mutations in the pre-S/S region. We speculate that increased positive selective pressure on the HBV pre-S/S region may allow HBV to escape the immune pressure exerted by the host and lead to possible alterations in HBV physiology. Further insight into the pathophysiology of these coexisting characteristics will rely on investigating these potential mechanisms, including the alteration of T and B-cell epitopes, stimulation of innate immune responses, and possible alterations in HBV physiology.

Abbreviations

HBsAg
Hepatitis B surface antigen
CHB
Chronic hepatitis B
HCC
Hepatocellular carcinoma
ALT
Aspartate alanine aminotransferase
AST
Alanine aminotransferase
TB
Total bilirubinemia
cccDNA
Intrahepatic covalently closed circular DNA
LR
Low replicative phase
RA
Reactivation phase
Sn
Shannon entropy
dS

Number of synonymous substitutions per synonymous site

dN

Number of non-synonymous substitutions per nonsynonymous site

ORFs

Opening reading frames

RT

Reverse transcriptase

CP

Core promoter

Declarations

Availability of data and materials

The data could be obtained upon request to the corresponding authors.

Acknowledgements

We thank all individuals who participated in this study.

Funding

This work was supported by grants from the Department of the Education of Zhejiang province (grants Y201839729 to YP Wu), Jinhua Science and Technology Bureau (grants 2018-4-141 to YP Wu) and Yiwu Science and Technology Bureau (YP Wu).

Author information

Affiliations

Department of Clinical Laboratory, Fourth Affiliated Hospital of Zhejiang University School of Medicine, Jinhua, Zhejiang Province, China

Yingping Wu, Zhiqiang Zhu, Jianyong Wu, Wenzi Bi, Weigen Wu and Wei Xu

Contributions

ZQ Zhu performed the sample collection, clonal sequencing and analysis, data interpretation, and drafted the manuscript. JY Wu assisted with the clonal sequencing. WZ Bi assisted with the sequencing and data analysis. WG Wu and W Xu assisted with recruitment of RA CHB patients and clinical data collection. YP Wu was responsible for the study design and data interpretation, and was a major contributor to the manuscript editing and critical revision of the article. All authors read and approved the final manuscript. We thank XP Xia and KS Xu for their support. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Yingping Wu.

Ethics declarations

Ethics approval and consent to participate

The protocol used in this study has been approved by Zhejiang University and all patients have read and signed informed consent.

Consent for publication

All authors have given consent for publication.

Competing interests

The authors declare that they have no competing interests.

References

1. Bernal W, Auzinger G, Dhawan A, Wendon J. Acute liver failure. *Lancet*. 2020; 376:190–201.
2. Wong DK, Seto WK, Fung J, Ip P, Huang FY, Lai CL, Yuen MF. Reduction of hepatitis B surface antigen and covalently closed circular DNA by nucleos (t) ide analogues of different potency. *Clin Gastroenterol Hepatol*. 2013; 11(8):1004–1010.
3. Chuaypen N, Sripapun M, Praianantathavorn K, Payungporn S, Wisedopas N, Poovorawan Y, Tangkijvanich P. Kinetics of serum HBsAg and intrahepatic cccDNA during pegylated interferon therapy in patients with HBeAg-positive and HBeAg-negative chronic hepatitis B. *J Med Virol*. 2017; 89(1):130–138.
4. Martinot-Peignoux M, Asselah T, Marcellin P. HBsAg quantification to optimize treatment monitoring in chronic hepatitis B patients. *Liver Int*. 2015; 35 (1):82–90.
5. Belloni L, Allweiss L, Guerrieri F, Pediconi N, Volz T, Pollicino T, Petersen J, Raimondo G, Dandri M, Levrero M. IFN- α inhibits HBV transcription and replication in cell culture and in humanized mice by targeting the epigenetic regulation of the nuclear cccDNA minichromosome. *J Clin Invest*. 2012; 122:529–537.
6. Guidotti LG, Chisari FV. Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol*. 2006; 1:23–61.
7. Tseng TC, Kao JH. Clinical utility of quantitative HBsAg in natural history and nucleos(t)ide analogue treatment of chronic hepatitis B: new trick of old dog. *J Gastroenterol*. 2013; 48:13–21.
8. Martinot-Peignoux M, Lapalus M, Laouénan C, Lada O, Netto-Cardoso AC, Boyer N, Ripault MP, Carvalho-Filho R, Asselah T, Marcellin P. Prediction of disease reactivation in asymptomatic hepatitis B e antigen-negative chronic hepatitis B patients using baseline serum measurements of HBsAg and HBV-DNA. *J Clin Virol*. 2013; 58(2):401–407.

9. Yao CC, Hung CH, Hu TH, Lu SN, Wang JH, Lee CM, Chen CH. Incidence and predictors of HBV relapse after cessation of nucleoside analogues in HBeAg-negative patients with HBsAg ≤ 200 IU/ml. *Scientific reports*. 2017; 7: 1839.
10. Wong RJ, Nguyen MT, Trinh HN, Chan C, Huynh A, Ly MT, Nguyen HA, Nguyen KK, Torres S, Yang J, Liu B, Garcia RT, Bhuket T, Baden R, Levitt B, da Silveira E, Gish RG. Hepatitis B Surface Antigen Loss and Sustained Viral Suppression in Asian Chronic Hepatitis B Patients: A Community-Based Real World Study. *J Viral Hepat*. 2017; 24(12):1089–1097.
11. Yang Y, Gao J, Tan YT, Li HL, Wang J, Ma X, Zheng W, Shu XO, Xiang YB. Individual and combined effects of hepatitis B surface antigen level and viral load on liver cancer risk. *J Gastroenterol Hepatol*. 2018; 33(5):1131–1137.
12. Honer Zu Siederdissen C, Cornberg M. The role of HBsAg levels in the current management of chronic HBV infection. *Ann Gastroenterol*. 2014; 27:105–112.
13. Wang CC, Tseng KC, Hsieh TY, Tseng TC, Lin HH, Kao JH. Assessing the durability of entecavir-treated hepatitis B using quantitative HBsAg. *Am J Gastroenterol*. 2016; 111:1286–1294.
14. Cornberg M, Wai-Sun Wong V, Locarnini S, Brunetto M, L.A H. Janssen, Lik-Yuen Chan H. The role of quantitative hepatitis B surface antigen revisited. *Journal of Hepatology*. 2017; 66:398–411.
15. Min AD. Low-Level Viremia in Hepatitis B Patients on Antiviral Treatment: Can We Ignore It? *HEPATOLOGY*. 2017; 66(2):312–314.
16. Nathanson MH, Terrault N. Hepatitis B surface antigen loss: not all that we hoped it would be. *HEPATOLOGY*. 2016; 64:328–329.
17. Pollicino T, Amaddeo G, Restuccia A, Raffa G, Alibrandi A, Cutroneo G, Favaloro A, Maimone S, Squadrato G, Raimondo G. Impact of hepatitis B virus (HBV) preS/S genomic variability on HBV surface antigen and HBV DNA serum levels. *Hepatology*. 2012; 56:434–443.
18. Brouwer WP, Chan HL, Brunetto MR, Martinot-Peignoux M, Arends P, Cornberg M, Cherubini B, Thompson AJ, Liaw YF, Marcellin P, Janssen HL, Hansen BE. Good Practice in using HBsAg in Chronic Hepatitis B Study Group (GPs-CHB Study Group). Repeated measurements of hepatitis B surface antigen identify carriers of inactive HBV during long-term follow-up. *Clin Gastroenterol Hepatol*. 2016; 14(10):1481–1489.
19. Velay A, Jeulin H, Eschlmann M, Malvé B, Goehringer F, Bensenane M, Fripiat JP, Abraham P, Ismail AM, Murray JM, Combet C, Zoulim F, Bronowicki JP, Schvoerer E. Characterization of hepatitis B virus surface antigen variability and impact on HBs antigen clearance under nucleo(t)ide analogue therapy. *Journal of Viral Hepatitis*. 2016; 23:387–398.
20. Abe K, Thung SN, Wu HC, Tran TT, Le Hoang P, Truong KD, Inui A, Jang JJ, Su IJ. Pre-S2 deletion mutants of hepatitis B virus could have an important role in hepatocarcino genesis in Asian children. *Cancer Sci*. 2009; 100:2249–2254.
21. Liu SJ, Zhang HW, Gu CY, Yin JH, He YC, Xie JX, Cao GW. Associations between Hepatitis B Virus Mutations and the Risk of Hepatocellular Carcinoma: A Meta-Analysis. *J Natl Cancer Inst*. 2009; 101:1066–1082.
22. Fang ZL, Sabin CA, Dong BQ, Ge LY, Wei SC, Chen QY, Fang KX, Yang JY, Wang XY, Harrison TJ. HBV A1762T, G1764A mutations are a valuable biomarker for identifying a subset of male HBsAg carriers at

- extremely high risk of hepatocellular carcinoma: a prospective study. *Am J Gastroenterol.* 2008; 103:2254–2262.
23. Bayliss J, Lim L, Thompson AJ, Desmond P, Angus P, Locarnini S, Revill PA. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. *J Hepatol.* 2013; 59: 1022–1028.
 24. Ayres A, Yuen L, Jackson KM, Manoharan S, Glass A, Maley M, Yoo W, Hong SP, Kim SO, Luciani F, Bowden DS, Bayliss J, Levy MT, Locarnini SA. Short duration of lamivudine for prevention of HBV transmission in pregnancy: lack of potency and selection of resistance mutations. *J Viral Hepat.* 2014; 21:809–817.
 25. Servant-Delmas A, Mercier-Darty M, Ly TD, Wind F, Alloui C, Sureau C, Laperche S. Variable capacity of 13 hepatitis B virus surface antigen assays for the detection of HBsAg mutants in blood samples. *J Clin Virol.* 2012; 53(4): 338–345.
 26. Pollicino T, Amaddeo G, Restuccia A, Raffa G, Alibrandi A, Cutroneo G, Favaloro A, Maimone S, Squadrato G, Raimondo G. Impact of hepatitis B virus (HBV) preS/S genomic variability on HBV surface antigen and HBV DNA serum levels. *Hepatology.* 2012; 56(2): 434–443.
 27. Bayliss J, Yuen L, Rosenberg G, Wong D, Littlejohn M, Jackson K, Gaggar A, Kittrinos KM, Subramanian GM, Marcellin P, Buti M, Janssen HLA, Gane E, Sozzi V, Colledge D, Hammond R, Edwards R, Locarnini S, Thompson A, Revill PA. Deep sequencing shows that HBV basal core promoter and precore variants reduce the likelihood of HBsAg loss following tenofovir disoproxil fumarate therapy in HBeAg-positive chronic hepatitis B. *Gut.* 2017; 66(11):2013–2023.
 28. Zagordi O, Däumer M, Beisel C, Beerewinkel N. Read length versus depth of coverage for viral quasispecies reconstruction. *PLoS One.* 2012; 7(10):e47046.
 29. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013; 30(12):2725–2729.
 30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19:1639–1645.
 31. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Mol Biol Evol.* 2007; 24(8):1586–1591.
 32. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005; 22(4):1107–1118.

Tables

Table 1. Characteristics of the study cohort. Data are shown as the n (%) and median (IQR). ALT, alanine aminotransferase; AST, aspartate aminotransferase; HBsAg, hepatitis B surface antigen.

Characteristics	Total(n=28)
Sex, Male(%)	49(69.01)
Age, years, median (IQR) (range)	46(23-82)
ALT, U/L, median (IQR) (range)	29(8-48)
AST, U/L, median (IQR) (range)	26(17-47)
HBsAg, IU/mL, median (IQR) (range)	32.85(0-88.37)
HBV DNA, LogIU/mL, median (IQR) (range)	3.71(3.30-7.25)
HBV DNA>20000 IU/ML(%)	42.25

Table 2. Codons in the HBV pre-S/ S regions under positive selection pressure.

Positively selected sites based on naive empirical Bayes (NEB) analysis (*, p>95%; **, p>99%). $\omega > 1$, means that significantly positive selection is

for adaptive protein evolution.

Region	Codon(pre-S/S)	Posterior probability	Omega (\pm s.e.m.)
pre-S1	47 L	1.000**	4.221 \pm 0.830
	60 A	0.992**	4.208 \pm 1.029
	62 A	0.955*	3.984 \pm 0.575
	73 G	0.984*	3.402 \pm 0.592
	84 I	0.978*	3.435 \pm 0.525
	90 A	0.969*	4.132 \pm 1.120
	128 H	0.979*	3.959 \pm 1.284
pre-S2	135 R	0.963*	3.646 \pm 1.472
	138Q	0.953*	4.217 \pm 0.838
	139 L	0.950*	4.027 \pm 1.230
	141 P	1.000**	4.221 \pm 0.830
	177 N	0.977*	4.159 \pm 1.089
S	374 Y	0.998**	4.219 \pm 0.835
	387 M	0.988*	4.206 \pm 0.857
	221 T	0.991**	3.441 \pm 0.512
	300 I	0.957*	3.324 \pm 0.717

Figures

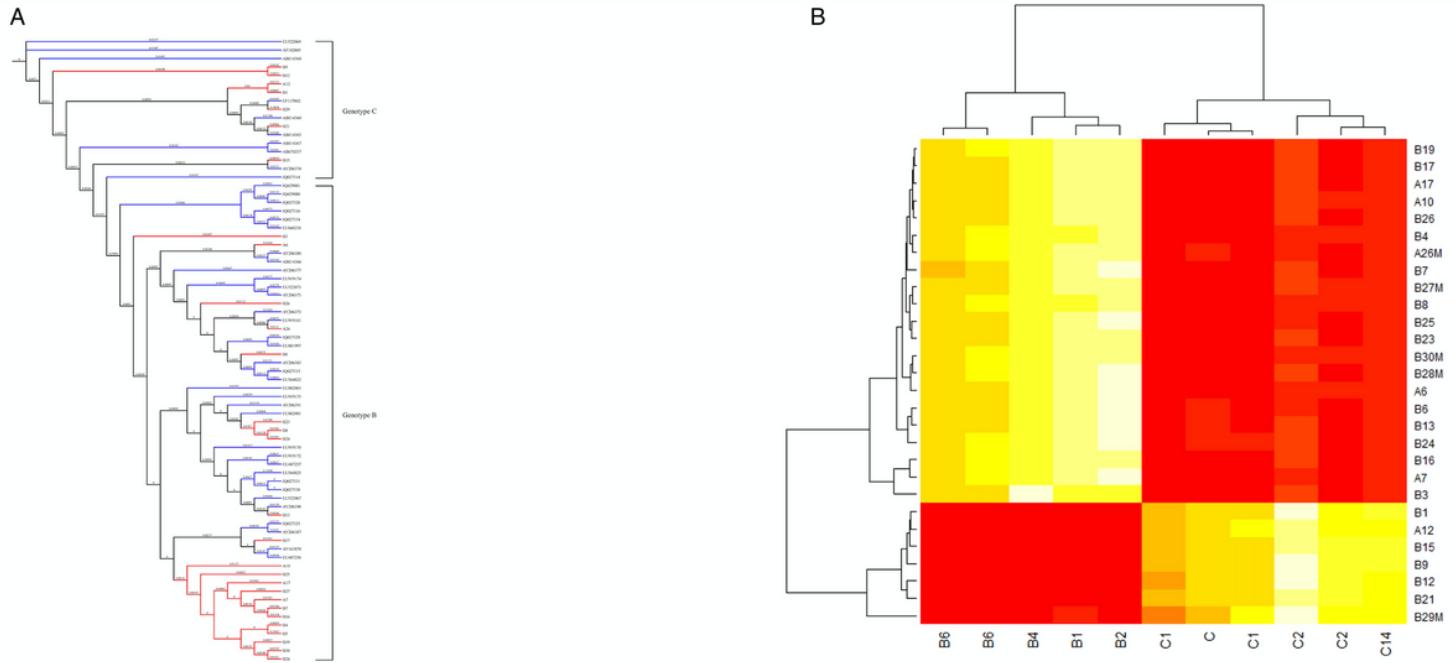


Figure 1

Phylogenetic tree of HBV isolates for genotyping and subgenotyping analysis. A. Phylogenetic trees were constructed for sequences from both the experimental group (red) and the reference sequences of HBV whole genome downloaded from NCBI (green) with a Bayesian method. Bootstrap values >90% are shown. From the phylogenetic analysis, 7 viral strains belong to genotype C and the others are genotype B. B. Subgenotyping analysis is clustered between HBV B2 and C2.

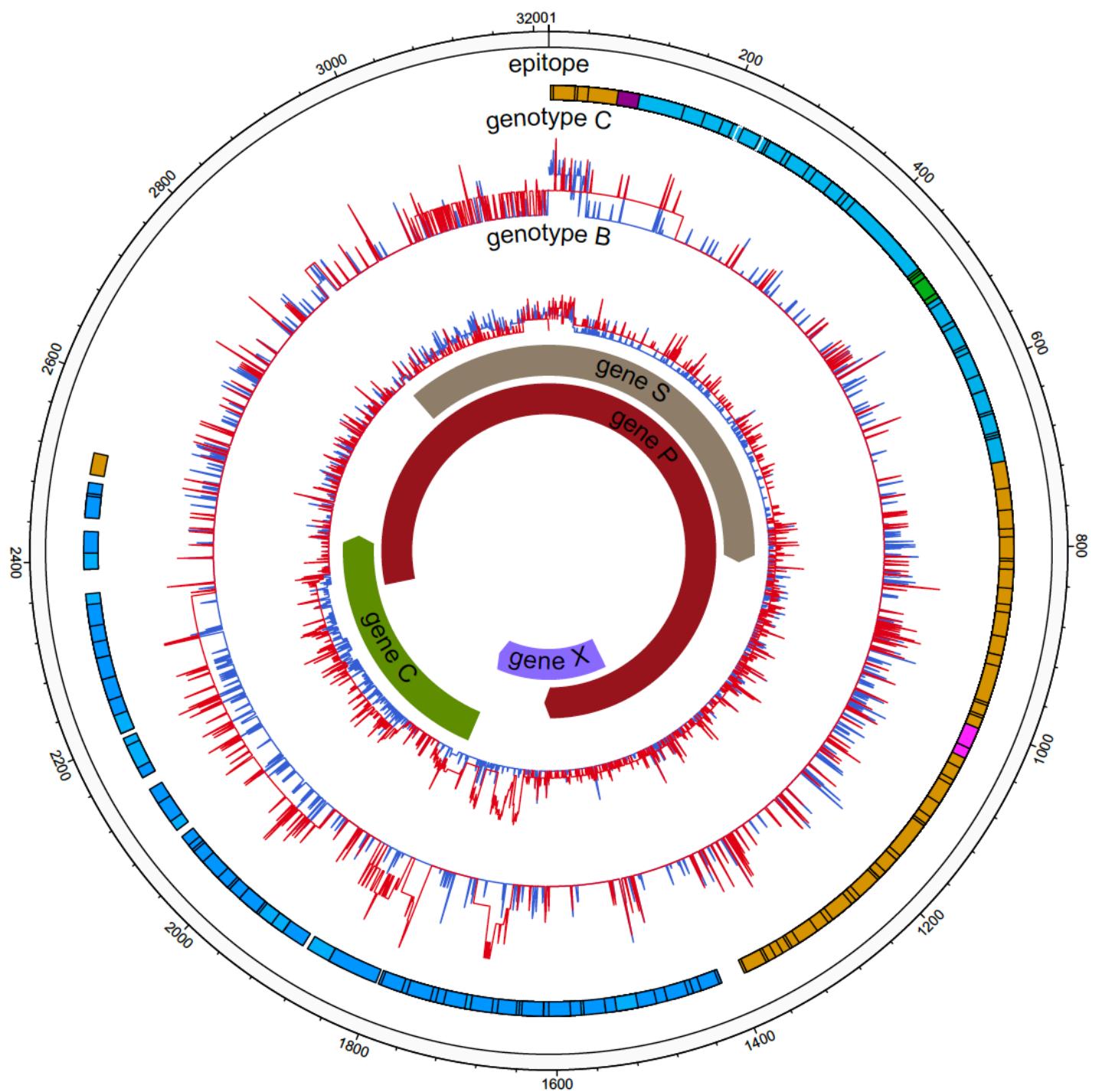


Figure 2

HBV nucleic acid complexity for the experimental group and control group (HCC). The colored bars indicate the complexity of each nucleotide for the experimental group (red lines) and control group (blue lines) for the full-length HBV genome. The insertions were discarded. Epitope distribution is as follows: B cell epitopes (purple), HLA I T cell epitopes (yellow), HLA II T cell epitopes (carnation), overlap of two types of epitopes (green), overlap of three types of epitopes (blue).

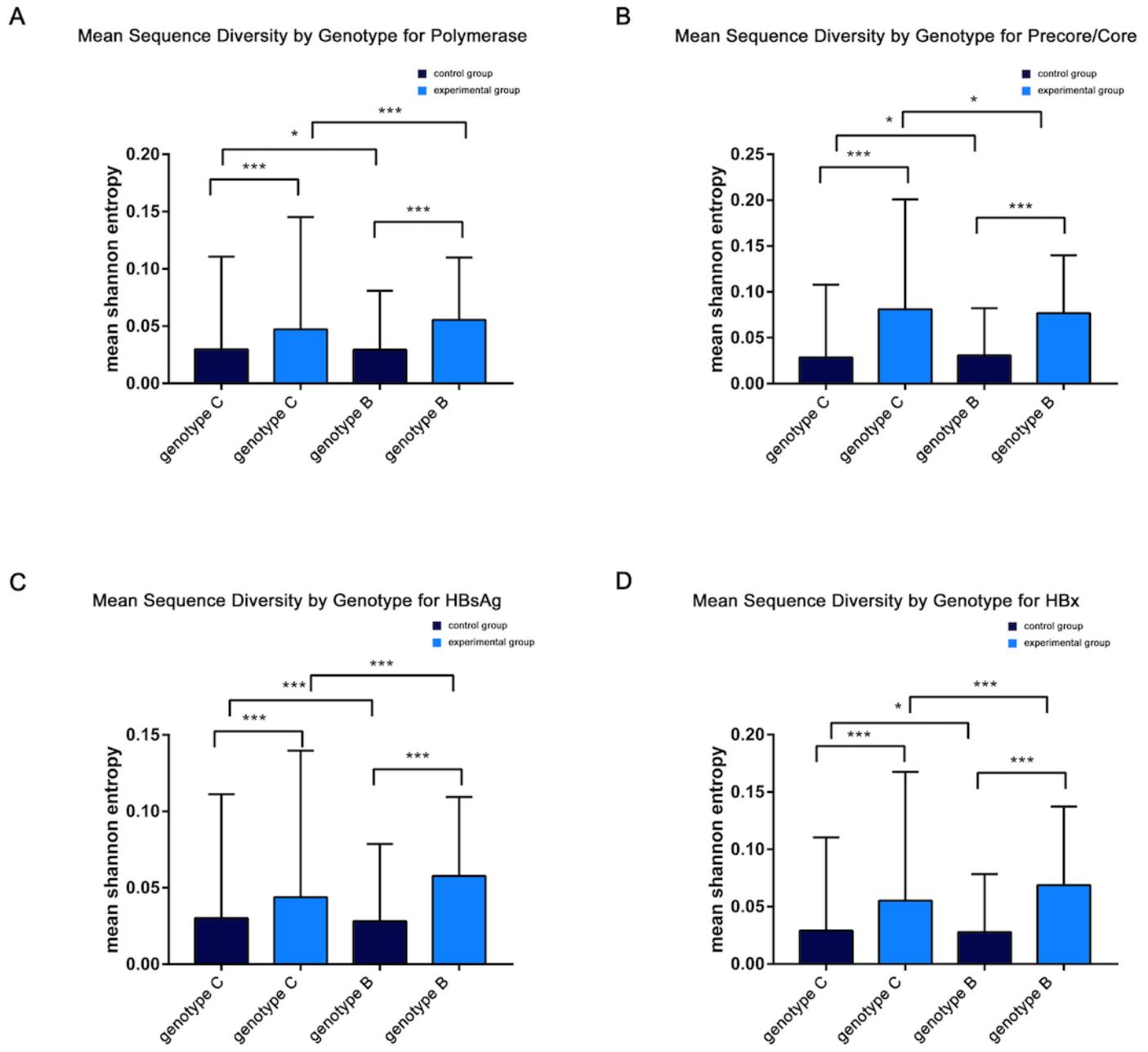


Figure 3

Mean sequence diversity across genotypes among baseline sequences identified by Shannon entropy in different regions of the HBV genome. Error bars indicate SEM; *** indicates $p < 0.001$; * indicates no significant difference. HBx, hepatitis B X; HBsAg,

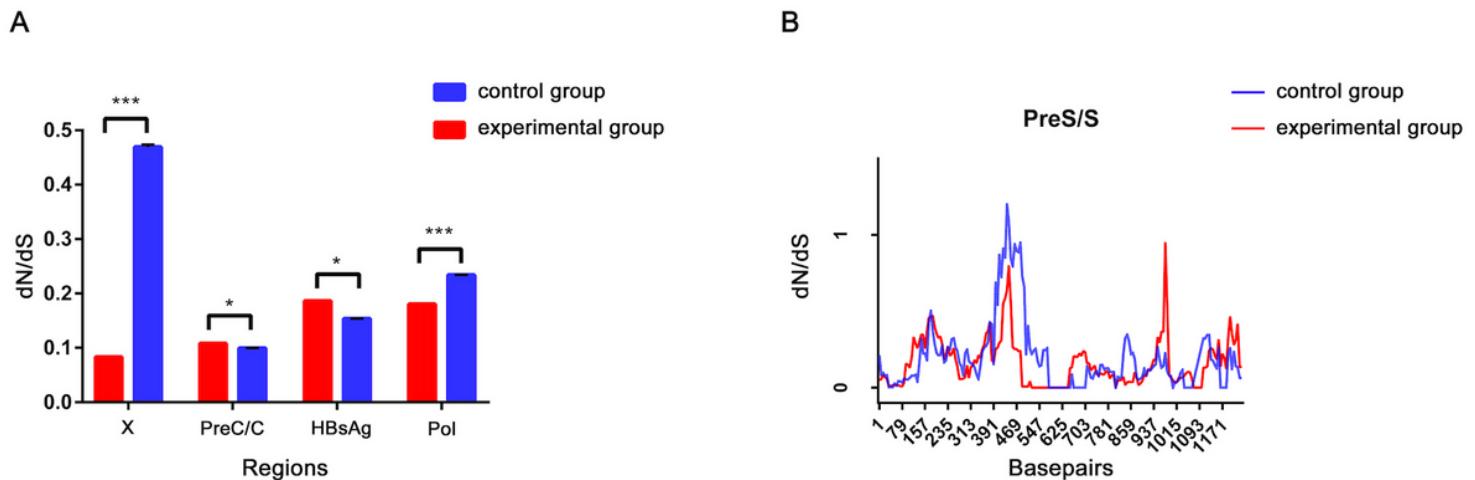


Figure 4

The dN/dS of the S gene was verified by a sliding window analysis using the KaKs calculator (window length, 57 nt; window step, 6 nt).

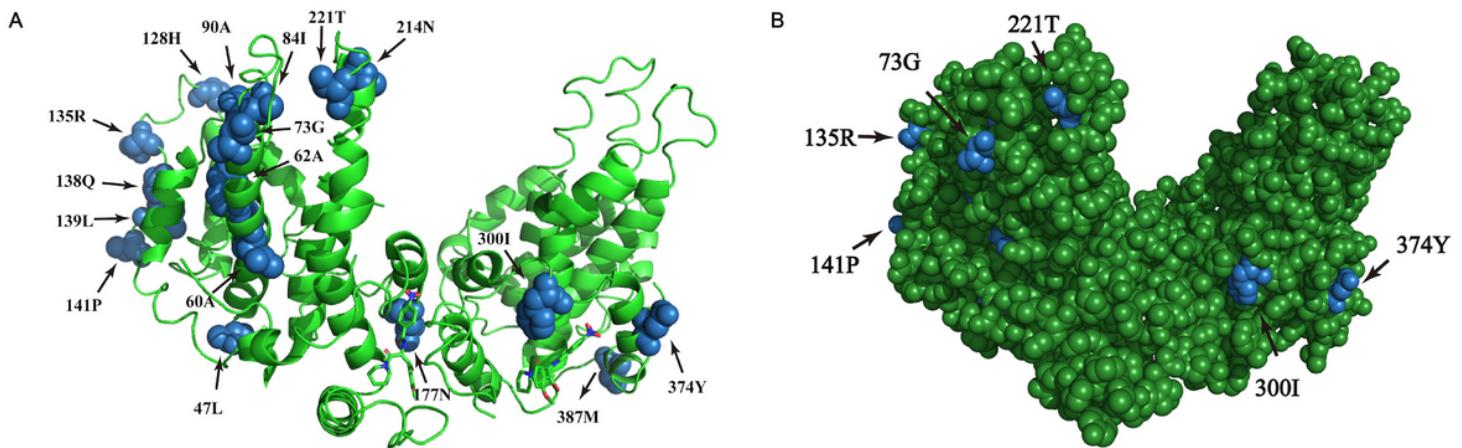


Figure 5

Homologous modeling analysis of sites of positive selection in HBsAg. The blue highlights the presence of point mutations within HBsAg and the green shows the HBsAg structure.