people worldwide have chronic HBV infections of HBV, with over 780,000 people dying annually due to the Hepatitis B disease (http://www.who.int/) and HCC being the third leading cause of cancer related deaths globally.

HBV has eight genotypes which are designated by capital letters of the alphabet from A to H and the nucleotide diversity of the intergroup is less than 8% [1]. Different HBV genotypes are clustered in different areas of the world. Genotype A is mainly present in Northwestern Europe, North America and Africa, genotype B and genotype C are mainly present in southeastern Asia, genotype D is mainly in Middle East, while genotypes E/F/G/H are predominant in Africa, South America, North America, and Central America respectively [1-4]. In China, genotypes B and C are the dominant infections, whereas those in western provinces of China are genotypes D and a C/D hybrid [5,6]. Previous studies showed that HBV genotypes differ by their pathogenic properties and also by epidemiological, virological and clinical variables [7]. There are also many disease-therapy resistant mutations associated with various genotypes [8,9].

Previous studies found some correlation between HBV genotypes and HBsAg serotypes [10,11]. According to the antigenicity difference of HBsAg, HBV was classified by nine serotypes. Generally, HBV isolates of genotype A or B were serotypes adw1 or adw2, genotype C isolates were serotypes adw2, adr or ayr, genotype D isolates were serotypes ayw2 or ayw3, and genotype E and F isolates were serotypes ayw4 and adw4 respectively [10,12].

HBeAg is one of the important clinical indexes of treatment response. For example, HBeAg-positive (HBeAg(+)) patients have a higher risk of LC or HCC [13]. Previous studies showed some correlation between specific gene mutations and HBeAg-negative (HBeAg(-)), meaning these could be treated as HBeAg(-) markers. Precore mutations G1896A, G1899A, Basal Core promoter (BCP) double mutations A1762T/G1764A, PreS1 mutation and especially PreS1 deletion all correlated with an increased risk of HCC. Accompanying the BCP double mutations A1762T/G1764A and PC mutation G1896A, the expression of HBeAg reduced significantly in the Chinese population [14-18]. In China, genotypes B and C are predominant and it has been found that the BCP double

**Keywords:** Hepatitis B virus (HBV); HBeAg(+/-); genotype B/C; evolutionary patterns

**1 Introduction**

Hepatitis B virus (HBV) is a member of the Hepadnaviridae family and has a (3200-bp) partially double-stranded circular DNA genome. This virus causes the chronic and acute disease Hepatitis B, such as liver cirrhosis (LC) and Hepatocellular carcinoma (HCC). More than 240 million
mutations and PC mutation were significantly different between genotypes B and C [19].

2 Datasets and Methods

2.1 Data processing

Complete HBV genomes were obtained from GenBank and the potential recombinant isolates detected by RDP3 [20,21]. The genotype of sequences was extracted from sequence annotation with self-developed PERL scripts. The HBV STAR program was used to check the genotype annotation of GenBank and classify the ambiguous genotyped sequences [22]. Sequences with inconsistent genotyped were abandoned, as were any sequences with insertions, deletions, internal stop codons or frameshifts in any ORF. While isolates collected in a province at one time, they were proposed as duplicated isolates for a subtype, so only one record was reserved as a representative of subtype.

The clean sequences were classified as HBeAg(+/-) according to absence or presence of the specific mutations of HBeAg(-) status described previously. BCP double mutation, Precore G1896A mutation, the deletion of ATG translational initiation codon of Precore, or a stop codon present in the preCore region were commonly found in HBeAg(-) [14-18]. Finally, all clean sequences were divided into four groups: B(+), B(-), C(+), C(-), named that B(+) represents HBeAg(+) of genotype B. Following the data processing described previously, we obtained 105, 131, 137, 148 sequences of B(+), B(-), C(+), C(-) respectively.

2.2 Sequence analysis

Multiple sequence alignment was performed using CLUSTALW, version 2.0.12 [23-25]. Group consensus sequences and group position frequency matrices were generated from the multiple alignments. ORF positions were referenced from AB602818 (genotype B) and DQ478899 (genotype C). The position whose maximum frequency was equal or greater than 0.95 was considered the conserved position.

2.3 Polymorphism parameter calculation

To evaluate the nucleotide diversity and polymorphism, sequences were mapped to group consensus sequence using NEEDLE, packaged of EMBOSS [26,27]. Single nucleotide polymorphism (SNP) was extracted from alignment by using self-developed PERL scripts. SNP ratio corresponds to the percentage of the number of substitutions divided by the gene length. The nucleotide diversity of genes was calculated as the average percentage of SNPs across the gene [28].

2.4 Evolutionary tree and evolutionary parameter calculation

The maximum likelihood phylogenetic tree was constructed for all sequences using MEGA, version 6.06 [29], with the maximum likelihood statistical method under the G+I model and with 1000 bootstrap replications. HBV genotype D sequence was used as the outgroup (Accession No. EU787441). The genetic distance (d) was calculated using MEGA6.0. Non-synonymous substitution rate (dN), synonymous substitution rate (dS) and the average ratio of non-synonymous to synonymous rates (dN/dS) were calculated using KaKs_Calculator2.0 with GMYN method [30,31].

2.5 Statistical analysis

The results of continuous variables were expressed as median and range, or as mean and standard deviation. Fisher’s two-tailed exact test and the chi-square test were used to compare qualitative data and Mann-Whitney U tests was used to compare continuous variables. All statistical calculations and charts were performed using R version 2.15. In all tests, statistical significance was set at p-value < 0.05.

3 Results

3.1 Phylogenetic tree

The maximum likelihood phylogenetic tree shows that all genotype B and C sequences were clustered separately. Genotype C was closer to outgroup genotype D than genotype B and also had a shorter phylogenetic tree length than genotype B (Fig. 1).

3.2 Conserved genome and polymorphism

HBV gene features and genome conserved overview are displayed in Fig. 2. HBV Genotype B was more conserved
Figure 1. The maximum likelihood phylogenetic tree constructed for genotype B and C sequences (bootstrap, 1000 replicates). The gray lines represent genotype B, black lines represent genotype C, and the marked red point is the outgroup of genotype D (EU787441).
Comparison of evolutionary patterns of different HBeAg status in HBV genotype B and C

557

than genotype C, especially in HBeAg(+) groups (Fig. 3). The non-conserved position count statistics of ORFs are shown in Table 1. The non-conserved position count of genes in B(+) were significantly lower than in C(+) except for PreC/Core. In genotype B the number was significantly lower in HBeAg(+) than in HBeAg(-), except for Precore. Whereas in genotype C, only S and Core were significantly different between HBeAg(+) and HBeAg(-). In HBeAg(-) groups, only P was significantly different in genotypes (Fig. 4 and Table 1).

3.3 Differential patterns according to HBeAg status

3.3.1 Nucleotide diversity of genes and gene overlapping

The nucleotide diversity of all genes was calculated for all groups. The diversity of S and preC were the lowest genes in all groups. Between genotype B groups, the mean diversity of P, S and Core of HBeAg(+) were 0.017, 0.006 and 0.015 respectively, and the mean diversity of P, S and
Core of HBeAg(-) were 0.019, 0.011 and 0.034 respectively (P < 0.05). The mean diversity of P, S and Core of HBeAg(+) were significantly lower than that of HBeAg(-). Among genotype C, mean diversity of X genes was significantly different, which was 0.021 in HBeAg (+), but 0.017 in HBeAg(-) (P < 0.01) (Table 2 and Fig. 5). The nucleotide diversity of gene overlapping regions was also calculated. Between genotype B groups, the nucleotide diversities of overlapping P and S, and P and C of HBeAg(+) were significantly lower than those of HBeAg(-). Non-Overlapping X and Non-Overlapping Core of HBeAg(+) were also significantly lower. Whereas between genotype C groups, the nucleotide diversities of overlapping P and X and Non-overlapping X of HBeAg(+) were significantly greater than HBeAg(-) (Table 3).

Figure 3. SNP distribution on whole genome. Group A, B, C, D means HBeAg(+) of genotype B, HBeAg(-) of genotype B, HBeAg(+) of genotype C, HBeAg(-) of genotype C respectively.
Comparison of evolutionary patterns of different HBeAg status in HBV genotype B and C

Table 1. Non-conserved Position counts of ORFs.

<table>
<thead>
<tr>
<th>ORF</th>
<th>ORF Length</th>
<th>Genotype B</th>
<th>Genotype C</th>
<th>HBeAg(+) P value</th>
<th>HBeAg(-) P value</th>
<th>HBeAg(+) P value</th>
<th>HBeAg(-) P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>2531</td>
<td>431</td>
<td>609</td>
<td>&lt; 0.001</td>
<td>0.096</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>RT</td>
<td>1032</td>
<td>158</td>
<td>217</td>
<td>0.001</td>
<td>0.163</td>
<td>0.028</td>
<td>0.206</td>
</tr>
<tr>
<td>preS1</td>
<td>357</td>
<td>69</td>
<td>108</td>
<td>0.001</td>
<td>0.417</td>
<td>0.003</td>
<td>0.628</td>
</tr>
<tr>
<td>preS2</td>
<td>165</td>
<td>31</td>
<td>53</td>
<td>0.008</td>
<td>0.303</td>
<td>0.004</td>
<td>0.206</td>
</tr>
<tr>
<td>S</td>
<td>678</td>
<td>63</td>
<td>105</td>
<td>0.001</td>
<td>0.012</td>
<td>0.014</td>
<td>0.113</td>
</tr>
<tr>
<td>X</td>
<td>465</td>
<td>62</td>
<td>98</td>
<td>0.002</td>
<td>0.311</td>
<td>&lt; 0.001</td>
<td>0.936</td>
</tr>
<tr>
<td>preC</td>
<td>87</td>
<td>4</td>
<td>10</td>
<td>0.162</td>
<td>0.103</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Core</td>
<td>552</td>
<td>112</td>
<td>197</td>
<td>&lt; 0.001</td>
<td>0.002</td>
<td>0.341</td>
<td>0.143</td>
</tr>
</tbody>
</table>

Table 2. Comparison of nucleotide diversity.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Genotype B</th>
<th>Genotype C</th>
<th>HBeAg(+) P value</th>
<th>HBeAg(-) P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>0.017 ± 0.054</td>
<td>0.019 ± 0.054</td>
<td>&lt; 0.001</td>
<td>0.026 ± 0.067</td>
</tr>
<tr>
<td>RT</td>
<td>0.014 ± 0.048</td>
<td>0.016 ± 0.049</td>
<td>0.008</td>
<td>0.027 ± 0.074</td>
</tr>
<tr>
<td>preS1</td>
<td>0.021 ± 0.060</td>
<td>0.018 ± 0.045</td>
<td>0.078</td>
<td>0.026 ± 0.062</td>
</tr>
<tr>
<td>preS2</td>
<td>0.019 ± 0.057</td>
<td>0.021 ± 0.051</td>
<td>0.166</td>
<td>0.029 ± 0.060</td>
</tr>
<tr>
<td>S</td>
<td>0.006 ± 0.029</td>
<td>0.011 ± 0.042</td>
<td>0.001</td>
<td>0.018 ± 0.061</td>
</tr>
<tr>
<td>X</td>
<td>0.012 ± 0.046</td>
<td>0.016 ± 0.055</td>
<td>0.166</td>
<td>0.021 ± 0.056</td>
</tr>
<tr>
<td>preC</td>
<td>0.005 ± 0.018</td>
<td>0.016 ± 0.059</td>
<td>0.145</td>
<td>0.004 ± 0.015</td>
</tr>
<tr>
<td>Core</td>
<td>0.015 ± 0.041</td>
<td>0.034 ± 0.074</td>
<td>&lt; 0.001</td>
<td>0.018 ± 0.044</td>
</tr>
</tbody>
</table>
3.3.2 Evolutionary patterns

Evolutionary patterns, including the genetic distance ($d$), Non-synonymous substitution rate ($dN$), synonymous substitution rate ($dS$) and the average ratio of non-synonymous to synonymous changes ($dN/dS$) were calculated for each group. In genotype B groups, the genetic distance ($d$) of all ORFs of HBeAg(+) were significantly shorter than that of HBeAg(-). In genotype C groups however, $d$ of all ORFs except preC/Core of HBeAg(+) were significantly longer than that of HBeAg(-).

$dN/dS$ values reflect the selection pressure of the genes. All of the ORFs had lower than 1 omega value. The median $dN$, $dS$ and $dN/dS$ of B(+) (ranged 0.007-0.037, 0.020-0.081 and 0.123-0.718, respectively) were significantly lower than those of B(-) (ranged 0.015-0.035, 0.026-0.095 and 0.245-0.769, respectively). In genotype C groups, $dN/dS$ values showed the same trend as those of HBeAg(+) (ranged 0.053-0.575) compared with HBeAg(-) (ranged 0.201-0.715). Whereas $dN$ and $dS$ showed a disparate trend with significantly different values of gene pairs (Fig. 6 and Table S1).
Comparison of evolutionary patterns of different HBeAg status in HBV genotype B and C

561

0.032 and 0.124-0.575, respectively). Again, in HBeAg(-) groups, the \( d \), \( dN \), \( dS \), \( dN/dS \) value showed no trend with significantly different values of gene pairs (Figure 6 and Table S1).

4 Discussion

The natural selective pressure and host immune pressure of HBV prompts a different evolutionary behavior along the genome [32]. In recent studies, Chinese patients with genotype C of HBV infection were more likely to suffer from chronic infection, which may progress to liver cirrhosis and Hepatocellular carcinoma [33]. In HBeAg(+) patients, HBV genotype B shows a better virological response to ADV therapy than genotype C [34]. Although the associations between HBV genotype B/C and response to lamivudine (LAM) therapy remains ambiguous, some studies have illustrated that YMDD mutation types show significant differences between HBV types B and C in Northern China [35].

In this study, we showed the different evolutionary patterns at two levels: genotype (B and C) and HBeAg serotypes (+ and -).

To compare the difference of evolutionary behaviors, we classified the sequences into four groups as described previously. The results suggested that genotype B was more conserved than genotype C at the genome level. In ORFs, genotype B presented significantly lower nucleotide diversity and had less SNP positions than genotype C, especially in HBeAg(+) subgroup. The X overlapped with BCP regions, contains negative regulatory elements (NRE) which play an important role in HBV replication.

### Table 3. Comparison of nucleotide diversity of gene overlapping.

<table>
<thead>
<tr>
<th>Genetic region</th>
<th>Region</th>
<th>Genotype B</th>
<th>Genotype C</th>
<th>HBeAg(+) vs. C</th>
<th>HBeAg(-) vs. C</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/preS1</td>
<td>2848-3204</td>
<td>0.021 ± 0.060</td>
<td>0.018 ± 0.045</td>
<td>0.078</td>
<td>0.026 ± 0.062</td>
</tr>
<tr>
<td>P/preS2</td>
<td>3205-3215, 1-154</td>
<td>0.019 ± 0.057</td>
<td>0.021 ± 0.051</td>
<td>0.166</td>
<td>0.029 ± 0.060</td>
</tr>
<tr>
<td>P/S</td>
<td>155-835</td>
<td>0.006 ± 0.029</td>
<td>0.011 ± 0.042</td>
<td>0.001</td>
<td>0.018 ± 0.061</td>
</tr>
<tr>
<td>X/preC</td>
<td>2307-2452</td>
<td>0.007 ± 0.030</td>
<td>0.017 ± 0.051</td>
<td>0.012</td>
<td>0.006 ± 0.021</td>
</tr>
<tr>
<td>Non-Overlapping P</td>
<td>1374-1623</td>
<td>0.010 ± 0.039</td>
<td>0.011 ± 0.047</td>
<td>0.486</td>
<td>0.020 ± 0.060</td>
</tr>
<tr>
<td>Non-Overlapping X</td>
<td>1814-1838</td>
<td>0.005 ± 0.014</td>
<td>0.013 ± 0.044</td>
<td>0.428</td>
<td>0.002 ± 0.004</td>
</tr>
<tr>
<td>Non-Overlapping C</td>
<td>1624-1813</td>
<td>0.015 ± 0.056</td>
<td>0.023 ± 0.065</td>
<td>0.016</td>
<td>0.023 ± 0.054</td>
</tr>
</tbody>
</table>

3.4 Differential patterns according to genotype B/C

3.4.1 Nucleotide diversity of genes and gene overlapping

In HBeAg(+) groups, the mean diversity of P, preS1, preS2 and X of genotype B were 0.017, 0.021, 0.019 and 0.012 respectively, and the mean diversity of P, preS1, preS2 and X of genotype C were 0.017, 0.021, 0.019 and 0.012 respectively. The mean diversity of P, preS1, preS2 and X of genotype B were significantly lower than that of genotype C. All of the gene nucleotide diversity between genotype B and C of HBeAg(-) were non-significant (Table 2 and Fig. 5).

The nucleotide diversity of gene overlapping of P and preS1/preS2 of genotype B were significantly lower than that of genotype C in HBeAg(+) groups. Nucleotide diversity of Non-overlapping X and P of genotype B were also significantly lower than that of genotype C in HBeAg(+) groups. Between HBeAg(-) groups, only non-overlapping X had significantly greater diversity of genotype B than that of genotype C (Table 3).

3.4.2 Evolutionary patterns

In HBeAg(+) groups, the \( d \) and \( dS \) value of genotype B (ranged 0.010-0.038 and 0.020-0.081, respectively) were significantly lower than those of genotype C (ranged 0.012-0.054 and 0.038-0.124, respectively). Whereas the \( dN \) and \( dN/dS \) value of genotype B (ranged 0.007-0.037 and 0.123-0.718, respectively) showed a different trend with significantly distinct values of genotype C (ranged 0.007-0.032 and 0.124-0.575, respectively). Again, in HBeAg(-) groups, the \( d \), \( dN \), \( dS \), \( dN/dS \) value showed no trend with significantly different values of gene pairs (Figure 6 and Table S1).
and cccDNA formation. S is the HBsAg coding gene which secretes HBsAg to stimulate the host immune response. In this study, X and S were the most conserved genes of the whole genome. According to traditional methods which evaluate selective pressures, a $dN/dS > 1$ or a $dN/dS$ which approaches 0 could be qualified as “positive or negative selection”, and as $dN/dS$ approaches 1 it is proposed as natural selection pressure. The $dN/dS$ of all ORFs from our study were less than 1. However, X and S were higher than other ORFs, and were close to 1. The genetic distance ($d$), $dN$ and $dS$ of X and S were extremely low. The data reveals that X and S are under natural selection pressure and will be at conserved status.

The evolutionary patterns were associated with HBeAg status. Genetic distance ($d$), $dN$, $dS$ and $dN/dS$ showed different evolutionary patterns between
HBeAg(+) and HBeAg(-). In genotype B, those values of HBeAg(+) were lower than those of HBeAg(-). The dS of HBeAg(+) was greater than that of HBeAg(-) in genotype C. This suggests that the synonymous substitution rate was higher in HBeAg(+) patients infected by HBV of genotype C, and more synonymous mutations will occur in genotype C. This could be one of the factors that lead to different mutation rates between genotypes B and C under antiviral therapy.

The nucleotide diversity of overlapping P and S/X/Core in genotype B was lower than that in genotype C, however this was not statistically significant. Whereas in HBeAg(+) subgroup, nucleotide diversity of non-overlapping X and C of genotype B was significantly lower. This means that the region which plays an important role in HBV replication may have lower nucleotide diversity, but this was not genotype specific.

The reverse transcription (RT) region is the substrate binding region of the nucleotide analogues (NAs). Drug resistant mutations, such as the YMDD mutation, has been reported to be associated with genotype[35]. In genotype B, the nucleotide diversity of HBeAg(+) was significantly lower than that of HBeAg(-), and the d, dN and dS of HBeAg(+) were also lower than that of HBeAg(-). While in HBeAg(+), the nucleotide diversity of genotype B was significantly lower than that of genotype C. However, in genotype C groups(C(+) and C(-)) and in HBeAg negative groups(B(-) and C(-)), the differences are more complicated and without obvious trends.

The observed evolutionary differences between genotypes B and C and the different HBeAg statuses could be used to improve the understanding of the Hepatitis B virus selection pressures and its mutations of immune escape.

Acknowledgements: This research was funded by Chongqing Research Program of Basic Research and Frontier Technology (No. cstc2014jcyjA10076).

Conflict of interest: authors declare nothing to disclose.

References


Supplemental Material: The online version of this article (DOI: 10.1515/biol-2015-0058) offers supplementary material.