Infection, Genetics and Evolution 97 (2022) 105184

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Analysis of entire hepatitis B virus genomes reveals reversion of mutations to wild type in natural infection, a 15 year follow-up study

Qin-Yan Chen [a], Hui-Hua Jia [a,b], Xue-Yan Wang [a], Yun-Liang Shi [a], Lu-Juan Zhang [a], Li-Ping Hu [a], Chao Wang [a], Xiang He [c], Tim J. Harrison [d], J. Brooks Jackson [e], Li Wu [f], Zhong-Liao Fang [a,*]

[a] Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control, Guangxi Key Laboratory for the Prevention and Control of Viral Hepatitis, Nanning, Guangxi 530028, China
[b] School of Preclinical Medicine, Guangxi Medical University, 22 ShuangYong Road, Nanning, Guangxi 530021, China
[c] Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou 511430, China
[d] Division of Medicine, University College London Medical School, London, UK
[e] Department of Pathology, Carver College of Medicine, University of Iowa, Iowa City, USA
[f] Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, USA

ARTICLE INFO

Keywords:
Hepatitis B virus
Evolution
Mutation
Reversion
Next-generation sequencing

ABSTRACT

It has been reported that some mutations in the genome of hepatitis B virus (HBV) may predict the outcome of the virus infection. However, evolutionary data derived from long-term longitudinal analysis of entire HBV genomes using next generation sequencing (NGS) remain rare. In this study, serum samples were collected from asymptomatic hepatitis B surface antigen (HBsAg) carriers from a long-term prospective cohort. The entire HBV genome was amplified by polymerase chain reaction (PCR) and sequenced using NGS. Twenty-eight time series serum samples from nine subjects were successfully analysed. The Shannon entropy (*Sn*) ranged from 0 to 0.89, with a median value of 0.76, and the genetic diversity (*D*) ranged from 0 to 0.013, with a median value of 0.004. Intrahost HBV viral evolutionary rates ranged from 2.39E-04 to 3.11E-03. Double mutations at nt1762(A → T) and 1764(G → A) and a stop mutation at nt1896(G → A) were seen in all sequences from subject BO129 in 2007. However, in 2019, most sequences were wild type at these positions. Deletions between nt 2920–3040 were seen in all sequences from subject TS115 in 2007 and 2013 but these were not present in 2004 or 2019. Some sequences from subject CC246 had predicted escape substitutions (T123N, G145R) in the surface protein in 2004, 2013 and 2019 but none of the sequences from 2007 had these changes. In conclusion, HBV mutations may revert to wild type in natural infection. Clinicians should be wary of predicting long-term prognoses on the basis of the presence of mutations.

1. Introduction

Hepatitis B virus (HBV) is the prototype virus of the family hepadnaviridae. It has a circular, partially double-stranded DNA genome of about 3200 nucleotides (nt) with four open reading frames (ORFs), the precore/core, polymerase, surface and X ORFs (Tiollais et al., 1985). This virus is characterized by a unique replication cycle that involves a reverse transcriptase (RT) step; the virus-encoded polymerase has RT, DNA-dependent DNA polymerase and protein priming activities. The polymerase lacks proofreading activity, leading to frequent genomic mutations, with the development of quasispecies in the individual host. Several genotypes have been recognized as a consequence of the long-term evolution of HBV (Lin and Kao, 2015; Revill et al., 2020).

HBV genomic mutations have been found in acutely and chronically infected patients and in all four open reading frames of the virus genome (Caligiuri et al., 2016). The most frequently occurring natural HBV variants are those with the precore stop and the basal core promotor (BCP) mutations. These mutations result in a reduction or abolition of the production of hepatitis B e antigen (HBeAg) and these mutations might be selected by the immune response of the host (Hadziyannis and Papatheodoridis, 2006).

According to an intergroup divergence of more than 8%, based on complete genomes, nine genotypes (designated genotypes A to I) and one putative genotype of HBV have been identified. The genotypes have

* Corresponding author at: Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control, 18 Jin Zhou Road, Nanning, Guangxi 530028, China.
E-mail address: zhongliaofang@hotmail.com (Z.-L. Fang).

https://doi.org/10.1016/j.meegid.2021.105184
Received 13 May 2021; Received in revised form 3 December 2021; Accepted 8 December 2021
Available online 11 December 2021

distinct geographic distributions (Okamoto et al., 1988; Norder et al., 1994; Kramvis, 2014). With between approximately 4 and 8% intra-group nucleotide differences across the complete genome and good bootstrap support, genotypes A-D, F and I have been further categorized into at least 49 subgenotypes (Zhang et al., 2016; Ren et al., 2019).

Accumulating evidence shows that mutations in the HBV genome are associated with certain disease manifestations, may affect the natural course of the infection, and confer resistance to antiviral agents (Baumert et al., 2005). HBV genotypes and variants may serve as markers to predict disease progression, as well as help physicians optimize individualized antiviral therapy in clinical practice (Lin and Kao, 2015).

Selection of the predominant viral strain is determined by factors such as the host immune response, viral replication fitness and exogenous pressures, such as antiviral therapy (Locarnini, 2005). Comparison of viral evolution over various periods of observation may provide more information to understand the accumulated sequence variations in the viral genome and the observed mutation rate over a long period, as well as viral pathogenesis. There have been a number of studies involving such comparisons (Osiowy et al., 2006; Betz-Stablein et al., 2016; Mina et al., 2017; Sakamoto et al., 2020). Currently, next generation sequencing (NGS) is likely the best approach to understand the nature of HBV evolution, diversity and quasispecies (Rybicka et al., 2016). However, there are few long-term longitudinal studies that compare full-length HBV genomes using NGS. In this study, we used NGS to determine the molecular evolution of HBV in asymptomatic HBsAg carriers from Guangxi, China, over a 15 year period.

2. Material and methods

2.1. Study subjects and ethic statement

Nine study subjects were selected from the Long An cohort (Fang et al., 2008). Eligible subjects were those whose serum samples were available at all four time points (2004, 2007, 2013 and 2019).All subjects were negative for human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) (Fang et al., 2008). None of them underwent anti-viral therapy before or during the study.

Informed consent in writing was obtained from each study subject. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki and has been approved by the Guangxi Institutional Review Board.

2.2. Serological testing

Sera were tested for markers (HBsAg, HBeAg/anti-HBe) of HBV infection, using enzyme immunoassays (EIA, Beijing Wantai Biopham Company Limited, Beijing, China). Alanine aminotransferase (ALT) levels were determined using a Reitman kit (DiaSys Diagnostic Systems (Shanghai), Shanghai, China).

2.3. Measurement of serum viral loads

Serum HBV DNA concentrations were quantified by real time PCR (Polymerase Chain Reaction, PCR) using commercial reagents (Sansure Biotech Inc., Hunan, China) in an ABI Prism 7500 sequence detection system (Applied Biosystems, Foster City, CA, California, USA), with a dynamic range of 1×10^2–5×10^9 IU/mL.

2.4. PCR for HBV genomic DNA

HBV genomic DNA was extracted from 200 μL of serum samples using QIAamp DNA Mini kits (QIAGEN GmbH, Hilden, Germany) and eluted in 50 μL of distilled water. The full-length HBV genome was amplified using PCR. The amplification protocol and primers P1 and P2 have been described previously (Günther et al., 1995). The amplification was performed for 40 of cycles with denaturation at 94 °C for 40s, annealing at 60 °C for 1.5 min, and elongation at 68 °C for 3 min, with an increment of 2 min after each 10 cycles, in a VeritiPro Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA). For samples which were not amplified using P1-P2 primers, nested PCR was carried out with 5 μLof the first round P1-P2 products in a 50 μL reaction using primers NPF1 (nt 1829–1851, 5'-ACCTCTGCCTAATCATCTCTTGT-3′) and NPR2 (nt 1822–1801, 5′- GTTGCATGGTGCTGGTGCGCAG-3′) and the same amplification protocol as the first round. Products from the second round were confirmed by electrophoresis through 1% agarose gels.

2.5. Library preparation and next-generation sequencing

PCR products from the second round PCR were sent to Delivectory Biosciences Inc. (Beijing, China). In brief, PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter. Beverly, Massachusetts) and quantified using Qubit dsDNA HS assay kits (Invitrogen, Carlsbad, CA, USA). Libraries of PCR products from each HBV whole genome were prepared using the Celero EZ DNA-Seq Library Preparation Kit (Tecan Genomics, Switzerland) and the sequences determined on a Noveseq sequencer (Illumina, San Diego, CA, USA), according to Illumina's protocol. Finally, fluorescent signals were analysed using the Noveseq control software and transformed to paired-end reads with 2*150 bps long sequences.

2.6. NGS data preprocessing and sample genotyping

Quality control and preprocessing of each sample's raw NGS short reads was performed by fastp v0.20.1 (Chen et al., 2018). The adapter sequences were removed and 15 bases of each read were trimmed from the 5'end. Any reads with an average quality score lower than 30, or length shorter than 50 nt, were filtered further. The remaining high quality reads from each sample were then mapped to a common reference sequence (accession no. X02763) using bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). After sorting and removing the duplications using Samtools v1.7 (Li et al., 2009), the consensus sequence was generated using CliqueSNV v1.5.3 (Knyazev et al., 2021). Consensus sequences of all samples were then multi-aligned with the reference sequences from HBVdb (Hayer et al., 2013) and three additional sequences, including FJ023664 of genotype I, AB486012 of genotype J, and AM117397 from Africa chimpanzees as the outgroup. A maximum-likelihood tree was then constructed using MEGA 7 (Kumar et al., 2016) and each sample was genotyped accordingly.

2.7. Haplotype construction and diversity analysis

The clean NGS reads from each sample were then realigned with the genotype specific references using bowtie2's very-sensitive-local strategy (Langmead and Salzberg, 2012), and deduplicated using Sambamba (Tarasov et al., 2016). The generated SAM files were used for haplotype reconstruction by implicating the CliqueSNV program (v1.5.3) with default settings (Knyazev et al., 2021). All haplotypes with a minimum abundance of 1% were used for downstream analysis, and each was considered as the genome sequence of a specific HBV variant. Generally, the viral quasispecies heterogeneity was evaluated by analyzing the genetic complexity, based on the number of different sequences present in the population. Shannon entropy and nucleotide diversity were then calculated, and the change over time was observed. Here Shannon entropy (S) was calculated by an inhouse script using the following formula: $S = -p_i \ln p_i$, where p_i is the frequency of each haplotype in the viral quasispecies population. The nucleotide diversity, D, was calculated as the mean pairwise genetic distance of the haplotypes obtained from MEGA X v10.1.8 (Kumar et al., 2018).

2.8. Estimation of the intra-host HBV evolutionary rate

For each subject, the viral longitudinal substitution process was

parameterised using an HKY substitution model (Hasegawa et al., 1985) as suggested by MEGA 7 (Kumar et al., 2016), and modelled among-site rate variation using a discretised C-distribution in the 'unconstrained' analyses, when estimating the intrahost viral evolutionary rates. Posterior estimates under the full probabilistic model were obtained using Markov Chain Monte Carlo sampling with a chain length of 100 million as implemented in BEAST v2.6.3 (Bouckaert et al., 2019). Convergence and mixing properties of the chains were inspected using Tracer v1.7.1 (Rambaut et al., 2018). Maximum clade credibility trees were summarized using the TreeAnnotator tool in BEAST and visualised in FigTree v 1.4.4 (Rambaut, 2009).

2.9. Statistical analysis

The data are presented as median (range). The medians were compared between groups using the Mann-Whitney test. The correlation analysis was carried using the Spearman test. A logarithmic transformation was applied to all viral loads prior to the analysis, to achieve an approximately normal distribution. All *P*-values were two-tailed and $P < 0.05$ was considered to be significant. All statistical analyses were performed using the SPSS software (ver.16.0; Chicago, IL, USA).

3. Results

3.1. General characteristics and genotypes

Nine subjects were included in this study, five males and four females. The average age in 2004 was 40.6 ± 7.2 years old. The ALT levels of these subjects were normal (6–40 U/L). Full length HBV genomes were successfully amplified from sera of three subjects at four sampling times, four subjects at three sampling times and two subjects at two sampling times.

Twenty-eight time series samples from the nine subjects were analysed successfully using NGS. On average, ~800,000 reads were maintained for each sample after quality filtering, corresponding to a mean coverage of 80,000 fold at each nucleotide site (Table 1). Consensus HBV sequences were constructed for all samples. As shown in Fig. 1, a maximum likelihood tree of these consensus sequences indicates that the 3 samples from 1 subject are of genotype B, the 4 samples from a second subject are of genotype I, and the remaining 21 samples from the other 7 subjects are of genotype C. Further analysis indicates that five subjects were infected with subgenotype C2. Two, one and one subjects were infected with subgenotypes C5, B4 and I1, respectively (Table 1 and Fig. 1).

3.2. Time series quasispecies and diversity

After realignment of the sequences from the samples with genotype-specific reference sequences, a median of 19 haplotypes with an abundance greater than 1% was obtained for each sample (Table S1). To examine the diversity of intrahost viral quasispecies, the Shannon entropy (*S*n) and mean pairwise genetic diversity were calculated for each sample. The *S*n ranged from 0 to 0.89, with a median value of 0.76, and the genetic diversity, *D*, ranged from 0 to 0.013, with a median value of 0.004. The *S*n values of most of the subjects' quasispecies increased initially and then declined (Fig. 2a, b); this change was seen for subjects CN149, CC246, TF006, TN122 and TS115. For these subjects, predominant viral strains were observed initially, then the quasispecies diverged and, finally, another strain became predominant. In two other subjects, BL71 and TX271, the *S*n values of the quasispecies increased throughout the observation period, indicating expansion of the quasispecies. In one subject, BO129, the *S*n value decreased from 2007 to 2019, indicating centralization of the quasispecies. Statistically, the Sn value was not associated with the viral load ($P = 0.561$).

The mean pairwise genetic diversity *D* of the subjects did not change dramatically, except for subject BO129. Statistics results also showed that the genetic diversity D was also not associated with viral load ($P = 0.630$), suggesting that the difference in the sequences of quasispecies in each subject was not significant.

3.3. Intrahost HBV viral evolutionary rates

There are three different evolutionary patterns in this study, with one example shown in Fig. 3a-c. In the first pattern, the most recent predominant strains had evolved from the previously predominant strains (such as in subjects TF006 and TX271). In the second pattern, all strains evolved and formed their own branch (subjects CC246, CN216 and TN122). In the third, the predominant strains gradually became minor strains while various minor strains expanded and became predominant (subjects BL71, BO129, CN149 and TS115).

The details of each subject's intrahost viral evolutionary rates are summarized in Table 1. The median value of the substitution rate of the subject infected with genotype B was 5.65E-4 (95% CI: 3.84E-4–7.48E-4) substitutions per site per year. The median value of the substitution rate of the subject infected with genotype I was 6.74E-4 (95% CI: 4.75E-4-8.87e-4) substitution per site per year. Four of the seven subjects

Table 1
Hepatitis B virus viral loads and evolutionary rates for each study subject.

Subjects	Sex	Age§	HBeAg	Genotypes	Viral Loads				Raw Reads				Evolutionary rates (Median) (95% CI, lower- upper)
					2004	2007	2013	2019	2004	2007	2013	2019	
BL71	F*	40	–	C2	79,276	1.84E+06	1.80E+03	6.98E+05	1,073,452	NA	NA	414,184	7.32E-04 (4.87E-04–9.78E-04)
BO129	F	53	–	C5	3.74E+04	1.38E+05	1.35E+04	3.52E+02	NA	464,655	NA	683,387	1.70E-03 (1.39E-03–2.06E-03)
CC246	F	37	+	C2	5.50E+07	1.11E+06	9.00E+05	4.40E+05	2,080,887	1,238,051	2,541,725	680,032	2.63E-04 (1.91E-04–3.36E-04)
CN149	M#	42	+	C2	1.50E+07	2.11E+06	1.42E+05	5.80E+04	1,008,374	775,412	NA	743,579	6.93E-04 (5.00E-04–8.85E-04)
CN216	M	35	–	C2	3.21E+05	1.56E+05	2.96E+05	2.00E+01	384,610	742,832	805,748	NA	1.01E-03 (7.61E-04–1.28E-03)
TF006	M	42	–	B4	24,438	7.25E+05	6.45E+05	3.70E+05	NA	428,880	1,303,502	979,415	5.65E-04 (3.84E-04–7.48E-04)
TN122	M	50	–	C2	13,468	2.46E+03	1.57E+04	4.93E+04	1,958,624	377,687	859,493	349,847	3.11E-03 (2.67E-03–3.58E-03)
TS115	F	35	–	I1	92,222	4.36E+04	9.00E+04	1.53E+05	2,768,401	257,362	764,086	1,292,869	6.74E-04 (4.75E-04–8.87E-04)
TX271	M	31	+	C5	1.63E+08	3.92E+07	1.39E+08	4.86E+02	1,380,425	1,236,421	1,558,516	NA	2.39E-04 (1.23E-04–3.60E-04)

M: male, * F: female, § The ages are those in 2004.

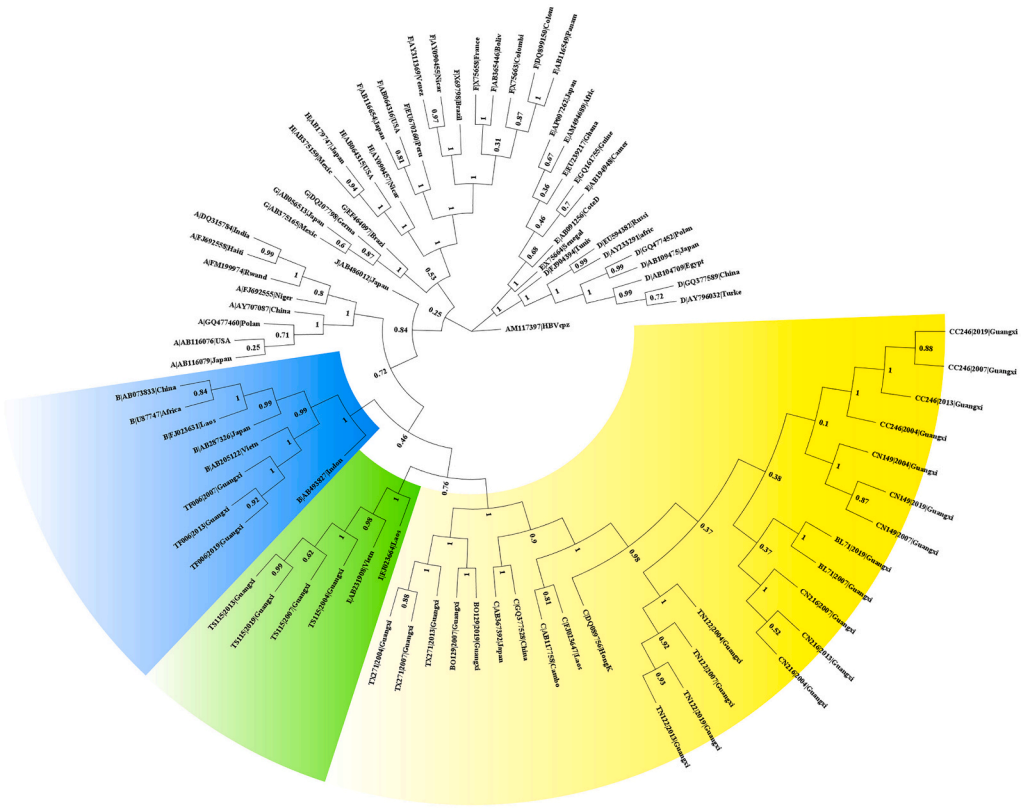


Fig. 1. Maximum-likelihood phylogeny of the consensus sequences from the temporal series of the nine subjects. The yellow coloured clade contains sequences from seven subjects infected with genotype C, CC246, CN149, BL71, CN216, TN122, BO129 and TX271. The green coloured clade contains sequences from subject TS115, infected with genotype I. The blue coloured clade contains sequences from subject of TF006, infected with genotype B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

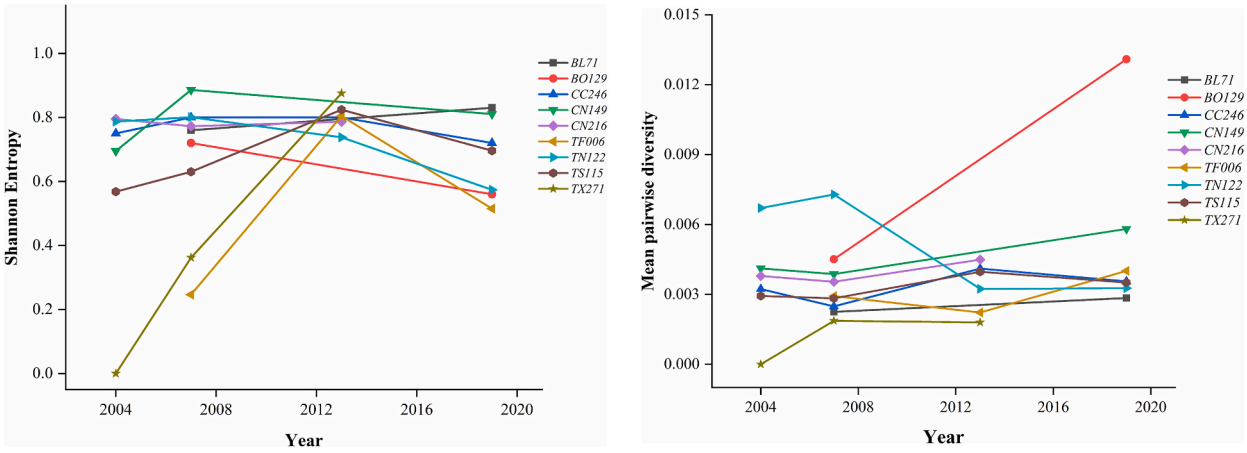


Fig. 2. Shannon entropy (Sn) (a) and genetic diversity (b) of the time series quasispecies from the 9 subjects.

infected with genotype C have similar intrahost viral evolutionary rates to those of the individuals infected with genotype B and genotype I, with the median value ranging from 2.39E-4 to 7.32e-4 substitutions per site per year. However, the other three subjects infected with genotype C (BO129, CN216, TN122) exhibited much higher intrahost viral evolutionary rates, with median values ranging from 1.01E-3 to 3.11E-4 substitutions per site per year. The viral loads of these three subjects were significantly lower than those of the other subjects ($Z = -3.054$, $P = 0.002$), suggesting that a high viral evolutionary rate is associated with a low viral load (Table 1). The difference in viral evolutionary rate between HBeAg positive and negative subjects is not significant ($Z = -1.807$, $P = 0.095$), suggesting that viral evolutionary rate is not associated with HBeAg status.

3.4. Characteristic mutations in the entire HBV genome

Mutations found in this study included point mutations and deletions detected by the Samtools mpileup algorithm and another inhouse script. Detailed information of variants with read depth greater than 1000 and mutation rate higher than 1.0% were shown in Table S2. Most of the point mutations are synonymous mutations. Four subjects had deletion mutations in the PreS1/S2/S region. Seven subjects had double mutations at nt1762 (A → T) and 1764 (G → A) in the basal core promoter at baseline (2004). In the remaining two subjects, the double mutations were present in 2007 and 2013. The PreC stop mutation at nt1896 (G → A) was seen in one subjects in 2004 and in four subjects during the follow up.

It seems that double mutations at nt1762 (A → T) and 1764 (G → A) are not associated HBeAg status. 22.2% (2/9) samples with these double

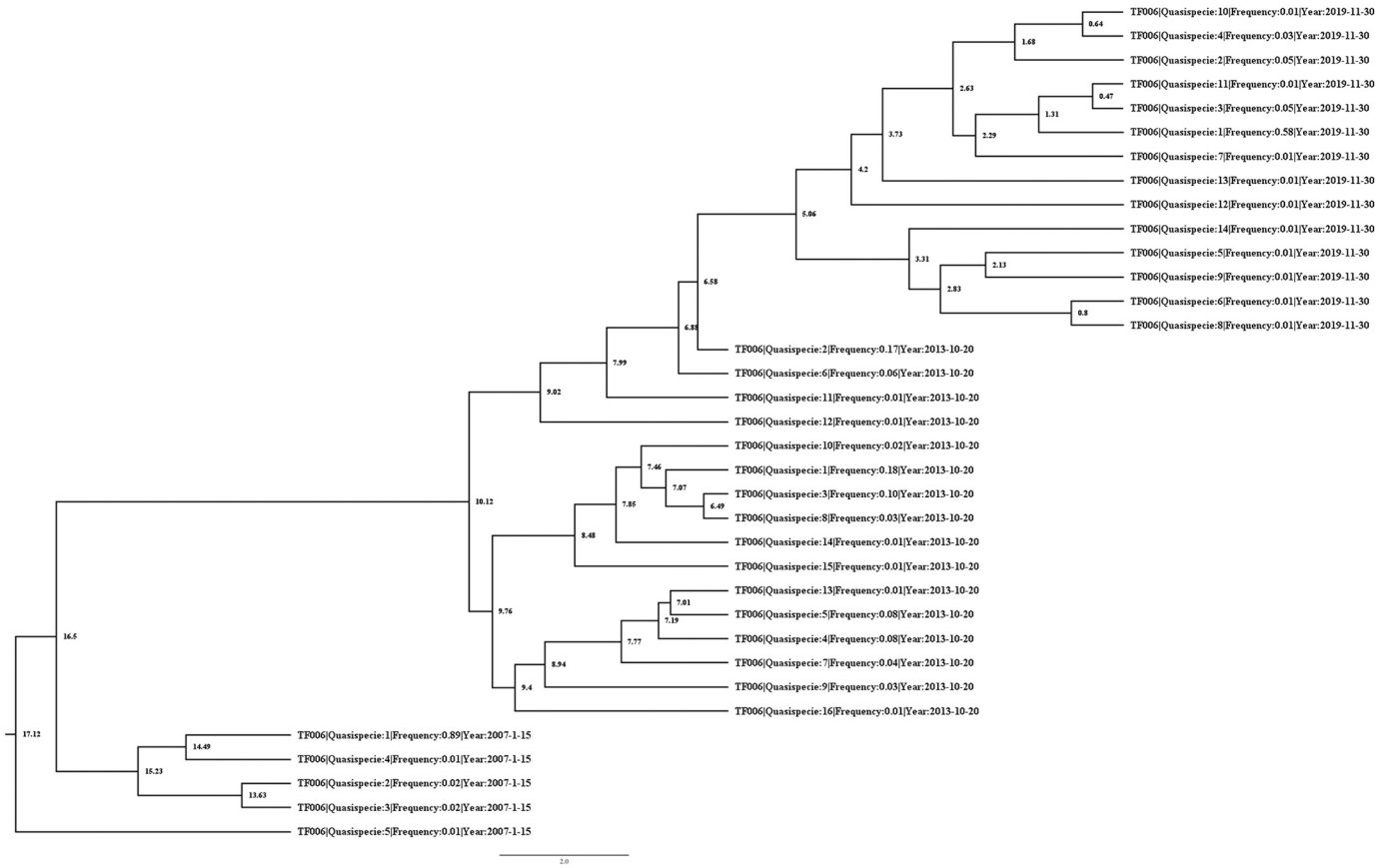


Fig. 3. Estimation of the hepatitis B virus intra-host evolutionary pattern and substitution rate.

mutations at baseline (2004) remain positive for HBeAg. One sample with wild type at core promoter remains positive for HBeAg, although the sequence evolved from wild type to double mutations. It is difficult to determine the association between PreC stop mutation (G → A at nt1896) and HBeAg status because four of the study subjects were negative for HBeAg before the mutation developed (Table 2).

There were some unique mutations in the HBV genomes from all study subjects, except for subject CN149. These mutations were located more often in the P gene and PreC/C gene than in the PreS1/S2/S gene and X gene (Fig. 4). Some are synonymous mutations, but none were antibody escaping or drug resistant mutations (Caligiuri 2016) (Table 3), suggesting that these unique mutations do not have clinical significance.

3.5. Reversion of mutations in the entire HBV genome

All haplotypes from subject BO129 in 2007 had the double mutations at nt1762(A → T) and 1764(G → A) and the stop mutation at nt1896(G → A), with both mutation rates greater than 99.0%. However, 43.8% (7/16) haplotypes with a total mutation rate of 30.0% were wild type at nt 1762 and 1764 while 37.5% (6/16) with a total mutation rate of 34.9% were wild type at nt1896 in 2019.

Subject TN122 had a deletion between nt 1–39. The rate of haplotypes was 81.5% (22/27), 89.3% (25/28), 96.2% (25/26) and 100% (17/17) sequences in 2004, 2007, 2013 and 2019, suggesting that the deleted viruses were becoming predominant and replacing the wild type. One haplotypes from subject CN216 in 2004 had a deletion between nt 43–54. However, none of the sequences had this deletion in 2007. In 2013, the subject had one sequence with a deletion between nt48-62.

Subject CC246 had point mutations (nt522C → A and nt587 G → A) in the S gene which are common mutations and led to antibody escape mutations (T123N, G145R) in the surface protein. The nt 522C → A mutation could be seen in 3.3% (1/30), 24.3% (9/37) and 28% (7/25) haplotypes in 2004, 2013 and 2019, respectively. The nt 587 G → A mutation was seen in 3.3% (1/30), 21.1% (8/37) and 24% (6/25) sequences in 2004, 2013 and 2019, respectively. However, none of 18 sequences in 2007 had either nt 522 C → A or nt 587 G → A mutations. Clearly, this subject had mutations at the beginning and end of the study but not in the middle of follow-up.

Subject TS115 had a deletion between nt 2920–3040 in all haplotypes in 2007 and 2013. However, none of the haplotypes from 2004 and 2019 had this deletion. A similar phenomenon was seen in subject CN149. One haplotypes from subject CN149 had a single mutation at nt1764 in 2007, but none of the sequences from 2004 or 2019 had this mutation. These two subjects had mutations in the middle of follow-up but not at the beginning or the end.

These data suggest that the appearance of HBV genomic mutations vary with time. Reversion of mutations in the entire HBV genome was not rare.

4. Discussion

To our knowledge, this is the first study to analyze long-term molecular evolution of full-length HBV genomes from asymptomatic carriers in China, using NGS. The major findings from this study are that HBV mutations may revert to wild type during natural infection. Both HBV genetic complexity and diversity vary with time. The HBV evolutionary rate and genetic diversity were associated with viral load but the Shannon entropy (*Sn*) was not. The strength of this study is that the

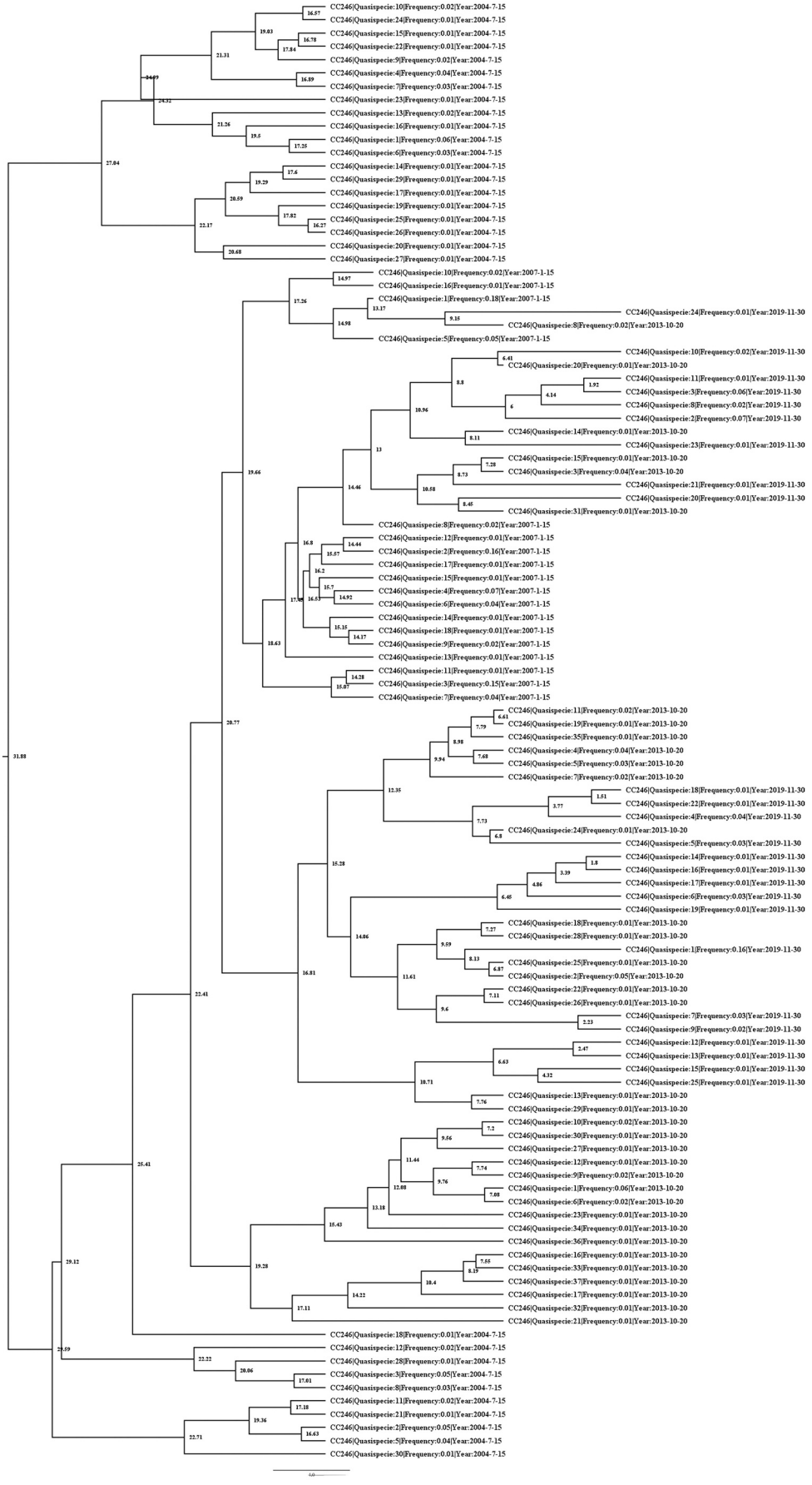


Fig. 3. (*continued*).

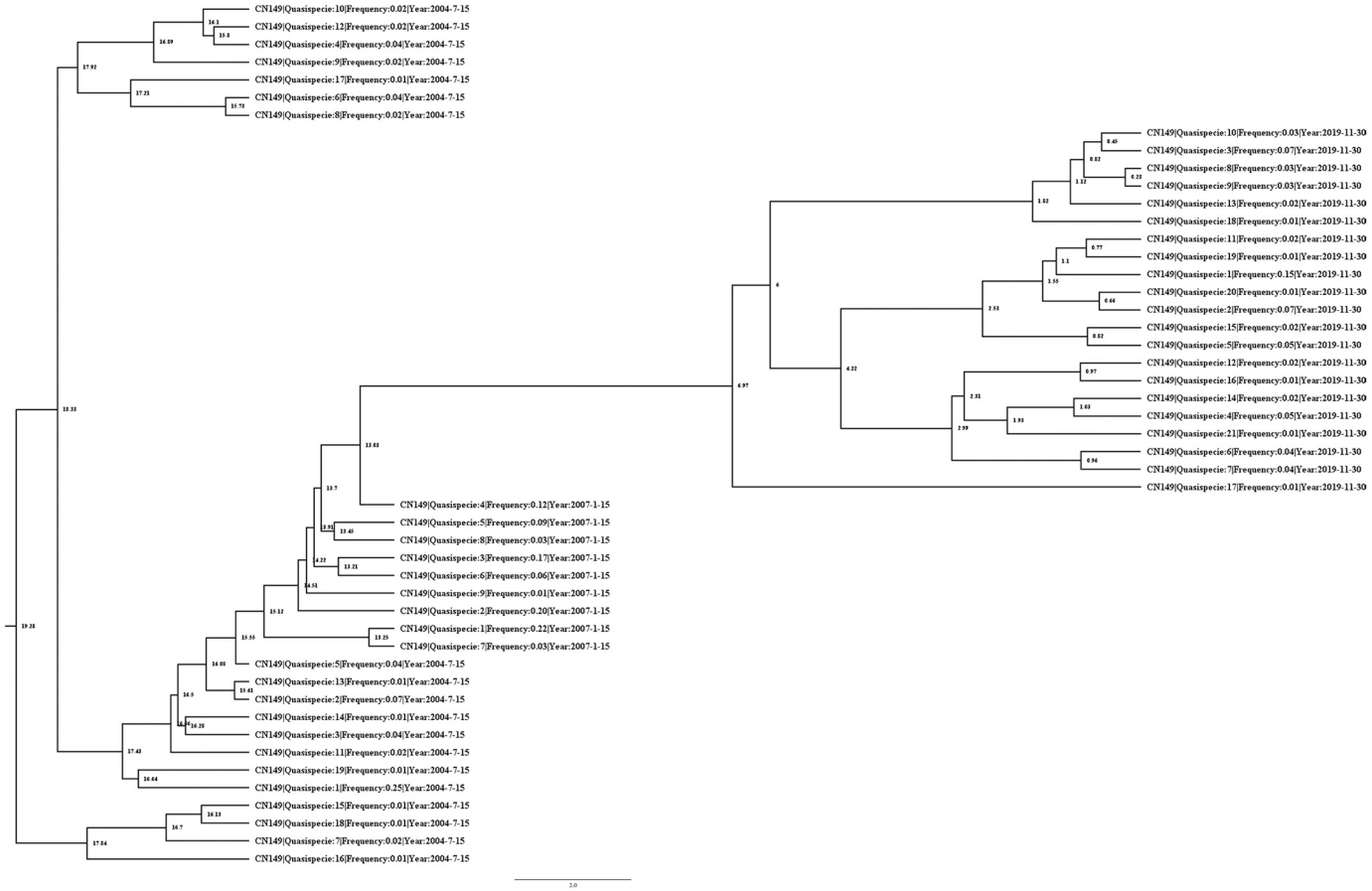


Fig. 3. (*continued*).

Table 2
Mutations in hepatitis B virus genome may have clinical significance.

Samples	2004		2007		2013		2019	
	PreS1/S2/S, nt1762,1764, nt1896	HBeAg	PreS1/S2/S, nt1762,1764, nt1896	HBeAg	PreS1/S2/S, nt1762,1764, nt1896	HBeAg	PreS1/S2/S, nt1762,1764, nt1896	HBeAg
BL71	1762A → T,1764G → A*	−	1762A → T,1764G → A (>99%)	−	1762A → T,1764G → A*	−	1762A → T,1764G → A (>99%)	−
BO129	1762A → T,1764G → A*	−	1762A → T,1764G → A (>99%) nt1896: G → A (>99%)	−	1762A → T,1764G → A*	−	1762A → T,1764G → A (70.0%) nt1896: G → A (65.1%)▽	−
CC246	1762A → T,1764G → A (>99%)	+	1762A → T,1764G → A (>99%)	−	1762A → T,1764G → A (>99%)	+	1762A → T,1764G → A (>99%)	+
CN149	nt 41–52 deletion (5.16%)# 1762A → T,1764G → A (94.5%)	+	1762A → T,1764G → A (98.1%)	+	1762A → T,1764G → A*	+	1762A → T,1764G → A (>99%)	−
CN216	nt 43–54 deletion (2.46%) 1762A → T,1764G → A (>99%) nt1896:G → A (32.1%)	−	nt 43–54 deletion (4.25%) 1762A → T,1764G → A (>99%) nt1896: G → A (21.55%)	−	nt 40–54 deletion (3.63%) 1762A → T,1764G → A (>99%) nt1896:G → A (>99%)	−	NA	−
TF006	WT in the three position*	−	1762A → T,1764G → A (>99%) nt1896: G → A (>99%)	−	1762A → T,1764G → A (>99%) nt1896: G → A (>99%)	−	1762A → T,1764G → A (>99%) nt1896: G → A (>99%)	−
TN122	nt38-54 deletion (93.1%) 1762A → T,1764G → A (98.5%)	−	nt38-54 deletion (92.4%) 1762A → T,1764G → A (96.1%)	−	nt38-54 deletion (97.6%) 1762A → T,1764G → A (>99%)	−	nt38-54 deletion (100.0%) 1762A → T,1764G → A (>99%)	−
TS115	1762A → T,1764G → A★	−	nt2907-2984 deletion (100.0%) 1762A → T,1764G → A (>99%) nt1896: G → A (5.0%)	−	nt2919-3041 deletion (100.0%) 1762A → T,1764G → A (>99%) nt1896: G → A (96.8%)	−	1762A → T,1764G → A (>99%) nt1896: G → A (65.8%)	−
TX271	WT in the three position	+	WT in the three position	+	1762A → T,1764G → A (>99%)	+	NA	−

*: Data from direct sequencing, #: Mutation frequency from the NGS data. ▽: Five of the 6 mutations (1896: G → A) are accompanied by core promoter double mutations (1762A → T, 1764G → A).

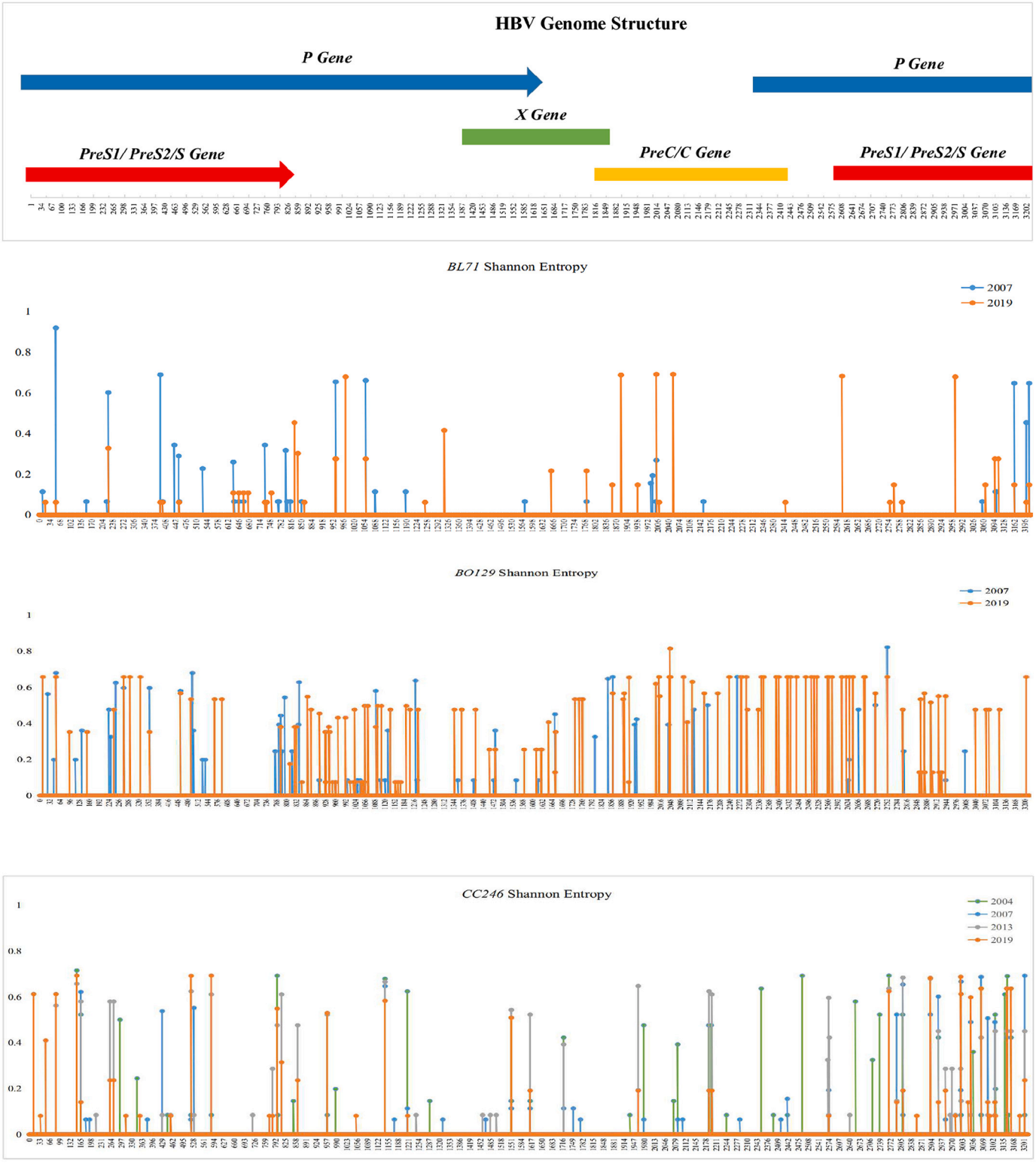
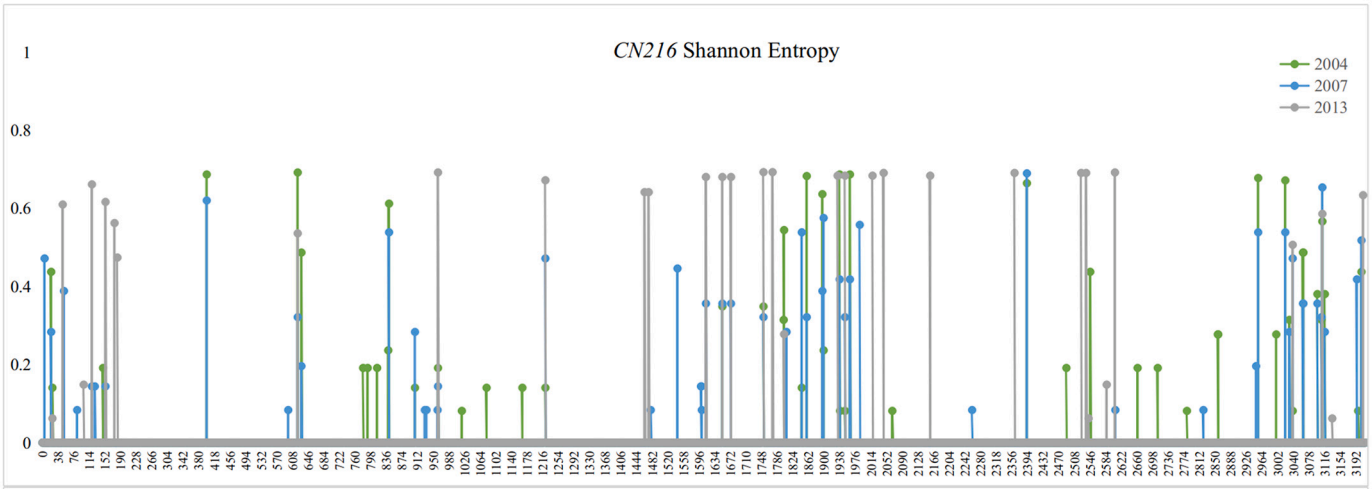
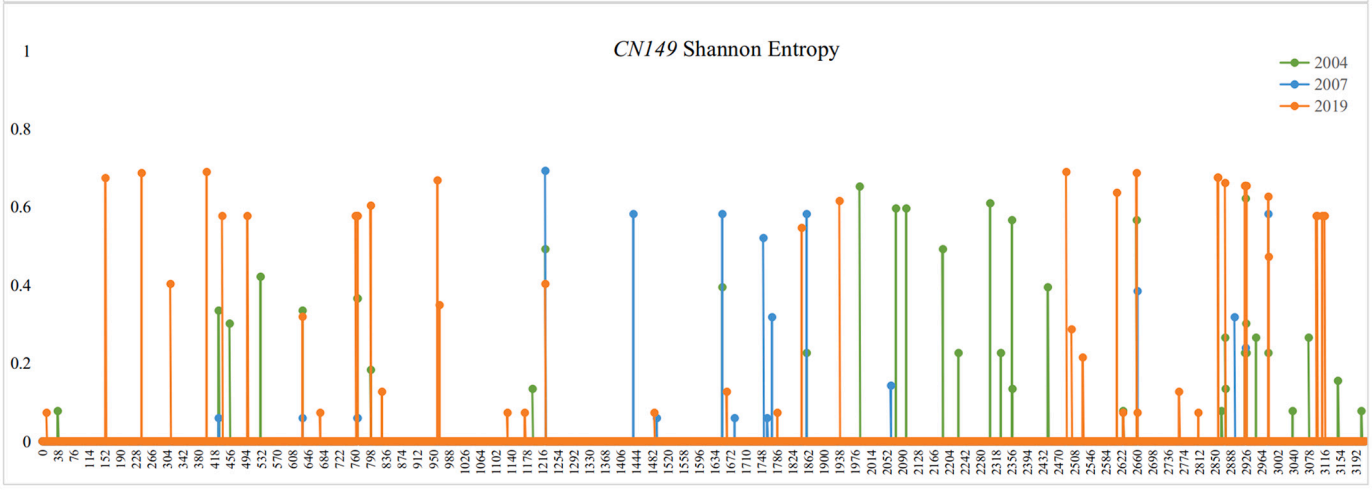


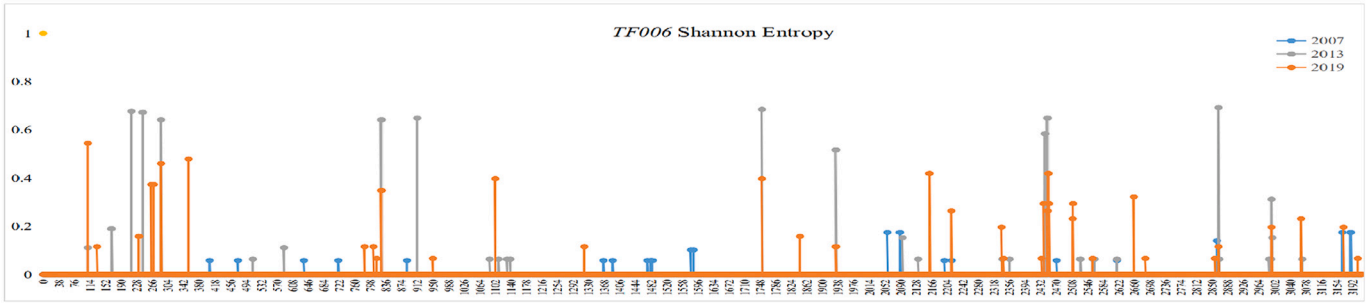
Fig. 4. Shannon entropy of hepatitis B virus intra-host viral quasispecies according to the genomic location.

study subjects were selected from a long-term prospective cohort, which allows us to analyze HBV molecular evolution over a 15-year period. The weakness is that the sample sizes were not sufficient for stratification analysis, such as analysis according to genotype.

The evolutionary origins of HBV and the timescale of its spread remain uncertain. Obtaining an accurate estimate of the rate of nucleotide substitution is the key to addressing the issue (Zhou and Holmes, 2007). However, the nucleotide substitution rate varies according to the region of the HBV genome. For example, the rate of substitution in regions of the HBV genome where ORFs overlap is 40% lower than in non-overlap regions and there is a significant difference in entropy between these regions (McNaughton et al., 2019). It has been







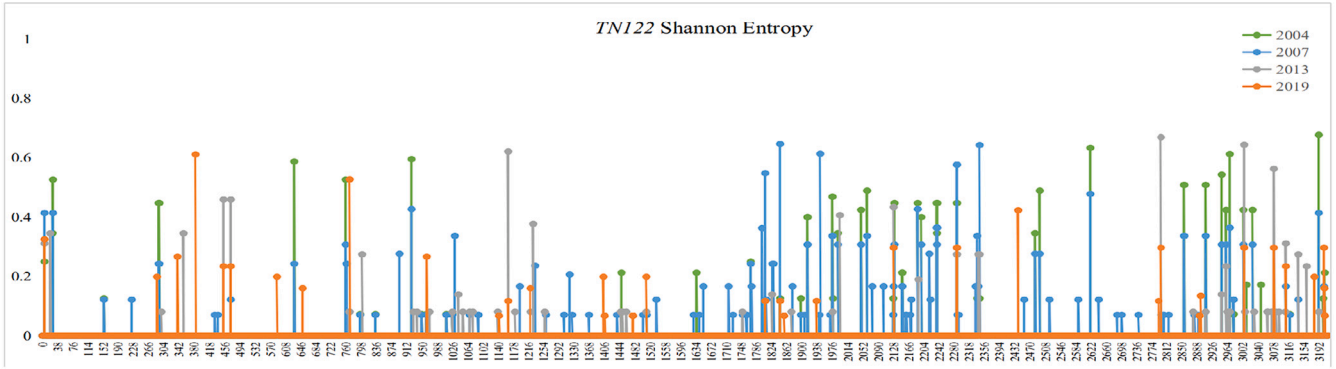


Fig. 4. (*continued*).

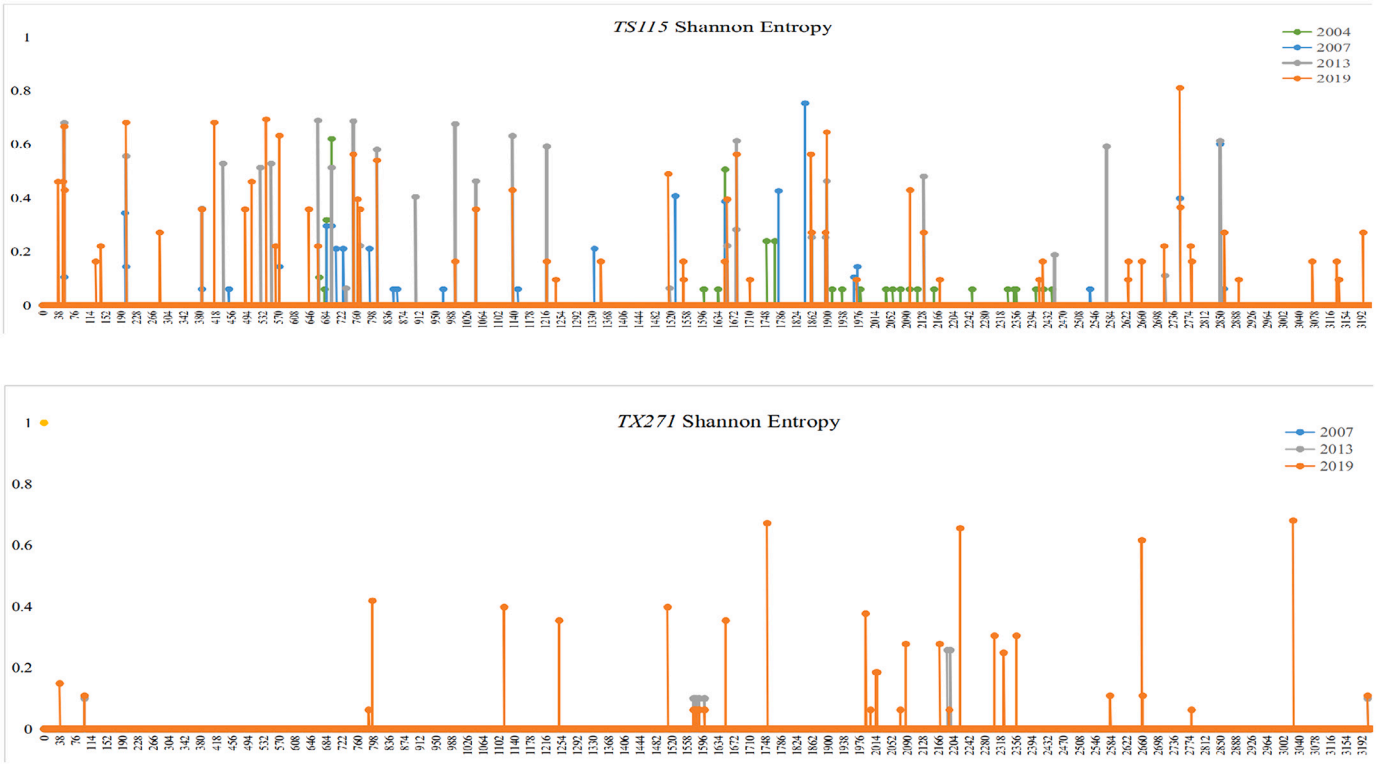


Fig. 4. (*continued*).

Table 3
Unique point mutations in the hepatitis B virus genome.

Subjects	BL71	BO129	CC246	CN149	CN216	TF006	TN122	TS115	TX271
Nucleotide changes (nt 1–3215#)	**S gene** 393 T → C (aa▼254F → S) **P gene** 393 T → C★ **C gene** 2170 T → C 2234A → C 2235G → A (aa141R → Q)▽ 2237G → C (aa142*E → Q) 2240A → G (aa143T → V)	**PrS2 gene** 154C → A (aa174N → K) **P gene** 154C → A (aa355H → N) 2429C → A 2430A → G 2431A → G (aa42 N → G) ☆ **C gene** 2048C → T 2404A → C 2049C → A (aa50: P → Y) 2240A → G (aa143: T → V) 2440C → A 2441A → G 2442A → G (aa206 Q → R)☆	**P gene** 959 T → C	0	**PrS1 gene** 3039G → A **P gene** 1230G → C (aa711G → A) 2582 T → C 3039G → A **C gene** 1970 T → C	**P gene** 954A → G 1257 T → G 1263G → T 2462G → C (aa52W → C) 2467A → G (aa55K → R) 2730C → T (aa150T → I)	**PreC/C gene** 1977C → A (aa55 S → N) 2063C → A (aa84 L → I) 2237A → C (aa141 R → S) 2239A → C (aa142 E → A)	**S gene** 348G → A 417 T → G **P gene** 348G → A 417 T → G 1424G → C (aa778V → L) **X gene** 1424G → C (aa17C → S)	**P gene** 1117C → A 1482G → T **X gene** 1482G → T

All of these mutations detected as changes from the baseline (2004) samples, except sample BL71, which was from 2019. # Nucleotide (nt) was numbered from 1 to 3215. ▼: aa: Amino acid.*: The position of aa was numbered according to its open reading frame. ★: Synonomous mutation. ▽ Two nt mutations caused one aa change. ☆Three nt mutations caused one aa change.

estimated from the results of Sanger sequencing over 25 years that the mean number of nucleotide substitutions/site/year for full-length HBV genomes in asymptomatic HBV carriers is 7.9×10^5 (Osiowy et al., 2006). Calculation of a three year evolutionary rate, based on NGS data of full-length HBV genome in asymptomatic HBV carriers, also suggested a rate of 4.42×10^5 substitution/site/year (Lin and Kao, 2015). Another 10 year analysis of full-length HBV genomes in asymptomatic HBV carriers reported a molecular evolutionary rate of 4.8×10^{-4} (Gauder et al., 2019). In this study, we found that intrahost HBV viral evolutionary rates are similar to that of that last analysis. Our data are based on long-term analysis (15 years) and NGS and, therefore, should be reliable.

It has been reported that the nucleotide substitution rate is inversely related to HBV DNA levels (Gauder et al., 2019). In contrast, we found in this study that a high evolutionary rate is associated with high viral loads. This issue needs to be addressed. It was reported that evolutionary

rates in the HBeAg-negative phase (anti-HBe-positive) are higher than in the HBeAg-positive phase (Hannoun et al., 2000). However, our results do not support this conclusion. This difference may be attributable to the small sample size.

Quasispecies complexity is a clinically relevant factor in the course HBV infection and the response to antiviral therapy (Xue et al., 2017; Trinks et al., 2020). Analysis of quasispecies complexity pre-treatment may be useful for managing patients with chronic hepatitis B (Homs et al., 2014). A high quasispecies complexity may result in a reduction of virus infectivity or lead to virus extinction in vitro (González-López et al., 2005). Analysis of HBV quasispecies complexity in pregnant women with high viral loads might be helpful to identify those whose babies may be at high risk of immunoprophylaxis failure (Xiao et al., 2020). Clearly, quasispecies complexity is important to understand the outcome of HBV infection (Lim et al., 2007). As an index of complexity, the Sn value may vary with the region of genome. It has been reported that the median based on the "a" determinant region of the surface protein was 0.0105 at the nucleotide level (Xiao et al., 2020); while that based on S and preC/C regions of the genome in chronic infection was 0.366 (Homs et al., 2014). The median Sn value in this study was higher. It is possible that the estimation in that study was based on a specific region of the genome while ours is based on the whole genome. However, this difference needs to be confirmed.

HBV replicates via an RNA intermediate, but its reverse transcriptase lacks the ability to proofread, leading to nucleotide misincorporation during genome replication (Simmonds and Midgley, 2005). This, in combination with a high replication rate, under the selective pressure from the administration of antiviral agents or an antibody response, leads to the development and selection of many variants during infection. It has been reported that antiviral therapy may result in reversion of precore/core promoter mutants to the wild type (Lin et al., 2004; Suzuki, 2003). In this study, we found that all sequences from subject BO129 in 2007 had BCP double mutations (nt1762T, 1764A) and the stop mutation (nt1896A). However, most of the sequences in 2019 were wild type at these positions. Another subject, TS115, had a deletion in PreS1 region of the genome in all sequences in 2007 and 2013. However, none of sequences from 2004 and 2019 had this deletion. These subjects were antiviral therapy-naive, suggesting that HBV mutations may have reverted to wild type during the natural course of infection without drug pressure. This reversion is possible because transmitted resistant HIV-1 also may revert to wild type in the absence of drug pressure, because of the reduced replication capacity of the resistant variants (Hofstra et al., 2013). It is also possible that the wild type strain persists at low abundance within the quasispecies when the viral populations are suppressed by immune pressure. When this suppression factor disappears, it will replicate more efficiently and dominate the viral quasispecies quickly. This needs to be confirmed. In this study, we do found that some wild type or mutated alleles were with a very low frequency usually less than 1%, which then could not be seen in the major haplotypes constructed by CliqueSNV (Data not shown).

The reversion of HBV mutations may reduce the predictive value of some mutations, in terms of clinical outcome, which could explain in part why some individuals with core promoter double mutations (nt1762T, 1764A) or PreS deletion mutations do not develop liver cancer, despite that these mutations have been suggested to be risk factors for liver cancer (Lin and Kao, 2015). These findings may help clinicians in considering these mutations when determining the prognosis of HBV infection.

It has been reported that that core promoter double mutations (nt1762T, 1764A) suppress, but do not abolish, the synthesis of HBeAg (Buckwold et al., 1996; Moriyama et al., 1996; Pang et al., 2004). In this study, it seems that the double mutations were not associated HBeAg status. It is also difficult in this study to determine the association between PreC stop mutation (nt1896 G → A) and HBeAg status, although the mutation could abolish the synthesis of HBeAg (Alexopoulou and Karayiannis, 2014). These findings may be attributed to small sample size.

Subject CC246 had antibody escape mutations in the S gene in 2004, 2013 and 2019. However, none of 18 sequences in 2007 had these mutations. It is not clear whether these mutants had been lost from the quasispecies or, perhaps, formed such a small proportion of the population that they were not detected by the deep sequencing.

In conclusion, the HBV genome evolutionary rate is high and associated with higher viral load. Both HBV genetic complexity and diversity vary with time. HBV mutations may revert to wild type without external pressure in natural infection, which possibly reduces the value of using the mutations to predict the clinical outcome.

Declaration of Competing Interest

The authors declare no competing interests..

Acknowledgements

We are indebted to staff members of Long An Center for Disease Prevention and Control and local hospitals in Long An county, Guangxi, who assisted in recruiting the study subjects, sample collection. This study was supported by the Global Health Research Seed Grant from the Carver College of Medicine at the University of Iowa, the National Natural Science Foundation of China (Grant No. 81860595 and 81703283), Guangxi Key Research and Development Project (Grant No. 2018AB59002) and Guangxi Natural Science Foundation (Grant No. 2017GXNSFBA198086).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105184.

References

Alexopoulou, A., Karayiannis, P., 2014. HBeAg negative variants and their role in the natural history of chronic hepatitis B virus infection. World J. Gastroenterol. 20 (24), 7644–7652.

Baumert, T.F., Barth, H., Blum, H.E., 2005. Genetic variants of hepatitis B virus and their clinical relevance. Minerva Gastroenterol. Dietol. 51 (1), 95–108.

Betz-Stablein, B.D., Töpfer, A., Littlejohn, M., Yuen, L., Colledge, D., Sozzi, V., et al., 2016. Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. J. Virol. 90 (16), 7171–7183.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15 (4), e1006650.

Buckwold, V.E., Xu, Z., Chen, M., Yen, T.S., Ou, J.H., 1996. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. J. Virol. 70, 5845–5851.

Caligiuri, P., Cerruti, R., Icardi, G., Bruzzone, B., 2016. Overview of hepatitis B virus mutations and their implications in the management of infection. World J. Gastroenterol. 22 (1), 145–154.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34 (17), i884–i890.

Fang, Z.L., Sabin, C.A., Dong, B.Q., Ge, L.Y., Wei, S.C., Chen, Q.Y., et al., 2008. 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. 103 (9), 2254–2262.

Gauder, C., Mojsiejczuk, L.N., Tadey, L., Mammana, L., Bouzas, M.B., Campos, R.H., et al., 2019. Role of viral load in hepatitis B virus evolution in persistently normal ALT chronically infected patients. Infect. Genet. Evol. 67, 17–22.

González-López, C., Gómez-Mariano, G., Escarmís, C., Domingo, E., 2005. Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe. Infect. Genet. Evol. 5, 366–374.

Günther, S., Li, B.C., Miska, S., Krüger, D.H., Meisel, H., Will, H., 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. J. Virol. 69 (9), 5437–5444.

Hadziyannis, S.J., Papatheodoridis, G.V., 2006. Hepatitis B e antigen-negative chronic hepatitis B: natural history and treatment. Semin. Liver Dis. 26 (2), 130–141.

Hannoun, C., Horal, P., Lindh, M., 2000. Long-term mutation rates in the hepatitis B virus genome. J. Gen. Virol. 81 (Pt 1), 75–83.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22 (2), 160–174.

Hayer, J., Jadeau, F., Deléage, G., Kay, A., Zoulim, F., Combet, C., 2013. HBVdb: a knowledge database for hepatitis B virus. Nucleic Acids Res. 41 (Database issue), D566–D570.
Hofstra, L.M., Nijhuis, M., Pingen, M., Mudrikova, T., Riezebos-Brilman, A., Simoons-Smit, A.M., et al., 2013. Evolution and viral characteristics of a long-term circulating resistant HIV-1 strain in a cluster of treatment-naive patients. J. Antimicrob. Chemother. 68 (6), 1246–1250.
Homs, M., Caballero, A., Gregori, J., Tabernero, D., Quer, J., Nieto, L., et al., 2014. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural evolution and on-treatment evolution. PLoS One 9 (11), e112306.
Knyazev, S., Tsyvina, V., Shankar, A., Melnyk, A., Artyomenko, A., Malygina, T., et al., 2021. Accurate assembly of minority viral haplotypes from next-generation sequencing through efficient noise reduction. Nucleic Acids Res. 49 (17), e102.
Kramvis, A., 2014. Genotypes and genetic variability of hepatitis B virus. Intervirology. 57 (3–4), 141–150.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547–1549.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. Nat. Methods 9 (4), 357–359.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25 (16), 2078–2079.
Lim, S.G., Cheng, Y., Guindon, S., Seet, B.L., Lee, L.Y., Hu, P., et al., 2007. Viral quasi-species evolution during hepatitis be antigen seroconversion. Gastroenterology. 133 (3), 951–958.
Lin, C.L., Kao, J.H., 2015. Hepatitis B virus genotypes and variants. Cold Spring Harb. Perspect. Med. 5 (5), a021436.
Lin, C.L., Liao, L.Y., Wang, C.S., Chen, P.J., Lai, M.Y., Chen, D.S., et al., 2004. Evolution of hepatitis B virus precore/basal core promoter gene in HBeAg-positive chronic hepatitis B patients receiving lamivudine therapy. Liver Int. 24 (1), 9–15.
Locarnini, S., 2005. Molecular virology and the development of resistant mutants: implications for therapy. Semin. Liver Dis. 25 (Suppl. 1), 9–19.
McNaughton, A.L., D'Arienzo, V., Ansari, M.A., Lumley, S.F., Littlejohn, M., Revill, P., et al., 2019. Insights from deep sequencing of the HBV genome-unique, tiny, and misunderstood. Gastroenterology. 156 (2), 384–399.
Mina, T., Amini-Bavil-Olyaee, S., Shirvani-Dastgerdi, E., Trovão, N.S., Van Ranst, M., Pourkarim, M.R., 2017. 15 year fulminant hepatitis B follow-up in Belgium: viral evolution and signature of demographic change. Infect. Genet. Evol. 49, 221–225.
Moriyama, K., Okamoto, H., Tsuda, F., Mayumi, M., 1996. Reduced precore transcription and enhanced core-pregenome transcription of hepatitis B virus DNA after replacement of the precore-core promoter with sequences associated with e antigen-seronegative persistent infections. Virology. 226, 269–280.
Norder, H., Couroucé, A.M., Magnius, L.O., 1994. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. Virology. 98 (2), 489–503.
Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miyakawa, Y., et al., 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J. Gen. Virol. 69 (Pt10), 2575–2583.
Osiowy, C., Giles, E., Tanaka, Y., Mizokami, M., Minuk, G.Y., 2006. Molecular evolution of hepatitis B virus over 25 years. J. Virol. 80 (21), 10307–10314.
Pang, A., Yuen, M.F., Yuan, H.J., Lai, C.L., Kwong, Y.L., 2004. Real-time quantification of hepatitis B virus core-promoter and pre-core mutants during hepatitis E antigen seroconversion. J. Hepatol. 40, 1008–1017.
Rambaut, A., 2009. FigTree version 1.3.1. http://tree.bio.ed.ac.uk.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian Phylogenetics using tracer 1.7. Syst. Biol. 67 (5), 901–904.
Ren, C.C., Chen, Q.Y., Wang, X.Y., Harrison, T.J., Yang, Q.L., Hu, L.P., et al., 2019. Novel subgenotype D11 of hepatitis B virus in NaPo County, Guangxi, bordering Vietnam. J. Gen. Virol. 100 (5), 828–837.
Revill, P.A., Tu, T., Netter, H.J., Yuen, L.K.W., Locarnini, S.A., Littlejohn, M., 2020. The evolution and clinical impact of hepatitis B virus genome diversity. Nat. Rev. Gastroenterol. Hepatol. 17 (10), 618–634.
Rybicka, M., Stalke, P., Bielawski, K.P., 2016. Current molecular methods for the detection of hepatitis B virus quasispecies. Rev. Med. Virol. 26 (5), 369–381.
Sakamoto, K., Umemura, T., Ito, K., Okumura, A., Joshita, S., Ota, M., et al., 2020. Virological factors associated with the occurrence of hepatitis B virus (HBV) reactivation in patients with resolved HBV infection analyzed through Ultradeep sequencing. J. Infect. Dis. 221 (3), 400–407.
Simmonds, P., Midgley, S., 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. 79 (24), 15467–15476.
Suzuki, F., 2003. Influence of the hepatitis B e antigen/anti-HBe status on the response to lamivudine. Intervirology. 46 (6), 339–343.
Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2016. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 31 (12), 2032–2034.
Tiollais, P., Pourcel, C., Dejean, A., 1985. The hepatitis B virus. Nature. 317 (6037), 489–495.
Trinks, J., Marciano, S., Esposito, I., Franco, A., Mascardi, M.F., Mendizabal, M., et al., 2020. The genetic variability of hepatitis B virus subgenotype F1b precore/core gene is related to the outcome of the acute infection. Virus Res. 277, 197840.
Xiao, Y., Sun, K., Duan, Z., Liu, Z., Li, Y., Yan, L., et al., 2020. Quasispecies characteristic in "a" determinant region is a potential predictor for the risk of immunoprophylaxis failure of mother-to-child-transmission of sub-genotype C2 hepatitis B virus: a prospective nested case-control study. Gut. 69 (5), 933–941.
Xue, Y., Wang, M.J., Yang, Z.T., Yu, D.M., Han, Y., Huang, D., et al., 2017. Clinical features and viral quasispecies characteristics associated with infection by the hepatitis B virus G145R immune escape mutant. Emerg. Microbes Infect. 6 (3), e15.
Zhang, Z.H., Wu, C.C., Chen, X.W., Li, X., Li, J., Lu, M.J., 2016. Genetic variation of hepatitis B virus and its significance for pathogenesis. World J. Gastroenterol. 22 (1), 126–144.
Zhou, Y., Holmes, C.E., 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. J. Mol. Evol. Actions 65 (2), 197–205.