Virus Research 273 (2019) 197762

Contents lists available at ScienceDirect

Virus Research

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

Characteristics and evolutionary history of hepatitis B virus quasi-subgenotype B3 in Southeast Asia

Jianmei Gao, Rongxia Zuo, Jinli Wang, Tao Shen*

Institute of Basic and Clinical Medicine, Yunnan Provincial Key Laboratory for Clinical Virology, Key Laboratory for Birth Defects and Genetic Diseases, the First People's Hospital of Yunnan Province, Kunming, China

ARTICLE INFO

Keywords:
Hepatitis B virus
Bayesian analysis
Nucleotide substitution rate
Time to most recent common ancestor

ABSTRACT

To analyze the hepatitis B virus (HBV) quasi-subgenotype B3 characters and molecular evolution in Southeast Asia, 411 serum samples with HBsAg positive were collected from Xishuangbanna, China. After DNA extraction, PCR amplification and sequencing, a total of 183 HBV full-length genomes were obtained. Phylogenetic analysis showed 139 stains (76.0%) were genotype B, 41 strains were genotype C (22.4%) and 3 strains were genotype I (1.6%). Among genotype B, 34 sequences were identified as quasi-subgenotype B3. Quasi-subgenotype B3 sequences from this study and quasi-subgenotype B3 sequences from the GenBank (total of 141 complete genome sequences) were grouped into quasi-subgenotype B3 (B3, formerly B5, Chinese B6 and B7-9). Sixteen peculiar nucleotides distributed in quasi-subgenotype B3 were identified, which were differ from B1, B2, B4 and B5(formerly B6) (nt93 T, nt100C, nt355 G, nt843 T, nt861C, nt912C, nt929 T, nt930 G, nt1023 T, nt1041 T, nt2651C, nt2693 T, nt2970C, nt3054A, nt3087A and nt3171 G). Then Evolutionary dynamics analysis of HBV quasi-subgenotype B3 was performed. The mean rate of nucleotide substitution for HBV quasi-subgenotype B3 was estimated to be around 5.556–5.660 $\times 10^{-4}$ substitutions/site/year. Estimated time to most recent ancestor of quasi-subgenotype B3 was around the 1847–1945(95%HPD), and Yunnan strains might be the parental strains. The Bayesian sky plot showed a steady spreading of HBV quasi-genotype B3 from early of 1940s to 90 s. In summary, HBV quasi-subgenotype B3 infection is prevalent in Southeast Asia based on the current reports and still with a high prevalence rate based on the evolutionary dynamics analysis. Current vaccine and nucleotide analogues might have effective prevention and treatment for HBV quasi-subgenotype B3 based on the rare clinically relevant mutation sites included in quasi-subgenotype B3.

1. Introduction

Hepatitis B virus (HBV) is a viral infection that attacks the liver and can cause both acute and chronic disease. According to the report of the World Health Organization, there were around 8.9 hundred thousand deaths resulted from hepatitis B in 2015. Hepatitis B virus belongs to the genus *Orthohepadnavirus* of the *Hepadnaviridae* family. To date, an intergroup divergence of greater than 7.5% over the entire genome sequence are used to separate strains into genotypes A–J, and an intragroup divergence of greater than 4% but less than 7.5% are used to identified several subgenotypes within genotypes, such as A (A1-A6), B (B1-B9), C(C1-C16), D(D1-D10), F(F1-F4) and I (I1-I2) (Hundie et al., 2017; Pourkarim et al., 2014; Ghosh et al., 2013). HBV genotypes and subgenotypes have distinct ethno-geographical distributions (Thedja et al., 2015) and been reported to be associated with different clinical and therapeutic outcomes (Tian and Jia, 2016; Ito et al., 2014; Li et al., 2017; Boglione et al., 2014). For example, many study showed that compared to HBV genotype C, HBV genotype B has earlier HBeAg seroconversion rates, higher HBsAg clearance rates, lower DNA level, slower liver damage and clinical progress, which means that infected with HBV genotype C might has stronger pathogenicity. Whilst comparing to subgenotype B1, infected with HBV B2 subgenotype were with more serious clinical characteristics, under long term of HBeAg positive status, and were more likely to lamivudine resistance variation. In addition, mutations T1809/ T1812 in the preC region and mutations T1762/A1764 in BCP region have been reported higher in HBV subgenotype A1 and C, respectively, which might affect the transcription of HBeAg precursor and incline development to liver cirrhosis and HCC. In our previous report, we investigated HBV infection status and prevalence genotypes in south-west part of China, Yunnan Province, which is contiguous to Burma, Laos and Vietnam, and is known to be one of the most ethnically diverse in China (Shen et al., 2009, 2015). In this

* Corresponding author.
E-mail address: ts902@126.com (T. Shen).

https://doi.org/10.1016/j.virusres.2019.197762
Received 18 March 2019; Received in revised form 17 September 2019; Accepted 17 September 2019
Available online 18 September 2019

report, we identified the HBV quasi-subgenotype B3 (formerly Chinese B6) in the Xishuangbanna, Yunnan Province. By reviewing literature and tracing the database, we find that the distribution of quasi-subgenotype B3 (including B3, formerly B5, Chinese B6, B7, B8 and B9) (Shi et al., 2012) was reported only in Indonesia, Philippines, Malaysia and Laos, but not in China except our findings. Then we collected reported HBV quasi-subgenotype B3 from GenBank by tracing the literature and performed dynamic analysis to date the history of HBV quasi-subgenotype B3 in south-east Asia.

2. Material and method

2.1. Subjects

411 serum samples including 174 Dai people, 71 Hanni people and 166 Han people aged 5–84 years old with HBsAg positive were collected from May 2012 to October 2012. All subjects were from outpatient department and inpatient department of Xishuangbanna Dai Autonomous Prefecture hospital, and they are indigenous inhabitants of Xishuangbanna, Yunnan Province. Informed consents were obtained from the participants. The protocol for the study was approved by the Ethical and Protocol Review Committee of the First People's Hospital of Yunnan Province.

2.2. HBV DNA amplification

HBV DNA was extracted from the serum samples using proteinase K digestion followed by phenol-chloroform extraction. Full genomes of HBV DNA were amplified by PCR. Primers for complete genomes amplification and sequencing were presented in Table 1. Amplification was performed in a 96-well cycler, 25 μL PCR mixture contained 200 μmol/L dNTP, 400 nmol/L of each primer, and 1U of Ex Taq polymerase (Takara BioTech). The full-length amplification was performed using the following cycles: 94°C pre-denature for 5 min, 30 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 3 min; and 72°C for 10 min as a final extension step. The amplicon were purified using a gel extraction kit and directly sequenced with BigDye Terminatorv3.1 and 3130 Genetic Analyzer (Applied Biosystems, Foster, USA).

2.3. Phylogenetic tree and characteristic analysis of HBV

The whole genome sequences were assembled from sequencing data, using BioEdit sequence alignment editor software, version 7.0.5.2. HBV sequences were aligned using Clustal X v1.83 and MEGA v5.05. Phylogenetic tree analysis was performed using the Kimura 2-parameter of evolutionary distance, and the topology was evaluated by bootstrap analysis (1000 replicates) using the neighbour-joining method. HBV genotype and subgenotype clustering were accomplished based on both the phylogenetic tree and the distances between and within the groups. Characteristics of sequences were analyzed by using DNAStar and BioEdit software packages. To identify the genotypes and subgenotypes, sequences from this study were compared with 9 reported HBV genotypes (A–I). Representative isolates of genotypes/subgenotypes A–I were retrieved from GenBank. Consensus sequences of 36 sequences of B1, 49 sequences of B2, 24 sequences of B4 and 38 sequences of B5, respectively, were used for the comparing with the quasi-subgenotype B3. The accession numbers are available from the authors on request.

2.4. Bayesian estimation of evolutionary rate and common ancestor date of HBV quasi-subgenotype B3

A total of 107 complete genome sequences of HBV quasi-subgenotype B3 were retrieved from the NCBI database (Supplementary Table 1). The nucleotide substitution rate and the time to most recent ancestor (tMRCA) were determined using the Bayesian Markov chain Monte Carlo (MCMC) approach implemented in BEAST v.1.7.1 (http://beast-mcmc.googlecode.com/files/STARBEAST.pdf). Sequences were divided into 5 sets of taxa based on the geographic origin, including quasi-subgenotype B3 data set, China (Yunnan) data set(CN-YN) (Shen et al., 2009), Indonesia data set (ID) (Utsumi et al., 2009; Mulyanto et al., 2010, 2011; Rinonce et al., 2013; Utsumi et al., 2015; Mulyanto et al., 2009; Okamoto et al., 1988; Nurainy et al., 2008; Thedja et al., 2011), Malaysia data set (MY) (Meldal et al., 2011; Chook et al., 2011; Suppiah et al., 2015), Philippine data set (PH) (Nagasaki et al., 2006; Sakamoto et al., 2006) and Laos data set (LA) (FJ023637 and FJ023638). The isolation year of sequences was introduced in a Bayesian coalescent analysis (tip dates method) in order to co-estimate an overall substitution rate and the tMRCA for HBV quasi-subgenotypes B3. In this study, the HYK substitution model was used with four gamma categories and SRD06 model. To model the population history, a Bayesian skyline plot (BSP) with 10 groups of intervals was used. To infer the ancestor location and viral migration events, a discrete phylogeographic analysis using a symmetric model was performed. As there is no consensus on the HBV substitution rate, we used relaxed molecular clock with initial values for the rate of substitution of 7.72×10^{-4} substitutions/site/year(high) (Zhou and Holmes, 2007), 1.0×10^{-5}substitutions/site/year (middle) (Torres et al., 2011) and 2.2×10^{-6} substitutions/site/year(low) (Paraskevis et al., 2013), respectively. Posterior distributions of parameters were obtained by MCMC analysis, which was run for 3.0×10^{9} steps in most cases, the parameter values were sampled every 30,000 steps. The program Tracer v1.7.1 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check for convergence and to determine whether sufficient mixing of

Table 1
PCR Primers for complete genomes amplification and sequencing.

Primer name	Primer sequence	Start-end	Primer length (bp)
HBV-AF	5 CTTTTTCACCTCTGCCTAATCA 3	1820-1841	22
HBV-AR	5 AGAGGTGAAAAAGTTGCATGGT 3	1811-1832	22
HBV-1R	5 GGGATGGGAATACAGGTGCAGTT 3	590-612	23
HBV-2F	5 CCAGGATCATCAACCACCAGCAC 3	485-507	23
HBV-3R	5 GGGAGTCCGCGTAAAGAGAGGT 3	1528-1549	22
HBV-5F	5 GCTTTAGGGCATGGACATTGACC 3	1891-1913	23
HBV-6R	5 AGACCAACCTCCCATGCTGTA 3	2842-2862	21
HBV-7F	5 ATTTTGCGGGTCACCATATTCT 3	2807-2828	22
HBV-8F	5 GCCGGTCTGGAGCGAAACTTATC 3	1304-1326	23
HBV-8R	5 TGGCCAGATTCATCAACTCACCC 3	2086-2108	23
HBV-10F	5 AAGAACTCCCTCGCCTCGCAGAC 3	2374-2396	23
HBV-10R	5 GGCGCTGCTGGCACTGTTGTC 3	3102-3122	21
HBV-11F	5 GCTCCTCTGCCGATCCATACTG 3	1252-1273	22
HBV-11R	5 GGCGAGGGAGTTCTTCTTCTAGG 3	2366-2388	23

the Markov chain sampler had been achieved in the posterior target distribution (effective sample size > 200 after a 10% burn-in were accepted). The maximum-clade-credibility tree was generated by the program TreeAnnotator v1.7.1 (available in BEAST) and the resulting tree file was visualized in the program FigTree 1.4.3.

2.5. Statistical analysis

Data was analyzed by using SPSS13.0 and were compared by chi-square and t-tests. A $P < 0.05$ was considered statistically significant.

3. Results

3.1. HBV genotype and sub-genotype classification

A total of 183 HBV full-length genomes were obtained by PCR amplification and sequencing. Phylogenetic analysis of all 183 HBV sequences showed that 139 strains (76.0%) were genotype B, 41 strains were genotype C and 3 strains were genotype I. Among genotype B isolates, subgenotype B2 was the major HBV genotype B (105/139, 75.540%), followed by quasi-subgenotype B3 (34/139, 24.460%). Among genotype C isolates, C1 was the major HBV genotype C (33/41, 80.488%), followed by C2 (4/41, 9.756%) and C5 (4/41,9.756%). In addition, quasi-subgenotype B3 was identified only in Hani people and Dai people, but not in Han people. Further analyses showed that the intragroup divergence of the entire genome sequences of quasi-subgenotype B3 was [mean ± SD (range)] 0.031 ± 0.002. The intergroup nucleotide divergences over the complete genome sequences between quasi-subgenotype B3 and the other subgenotypes were 0.055 ± 0.003 with B1, 0.045 ± 0.003 with B2, 0.044 ± 0.003 with B4, 0.051 ± 0.003 with B5 (formerly B6). These sequence data have been submitted to the GenBank databases under accession number of MK534554-MK534736. Addresses are as follows: GenBank: www.ncbi.nlm.nih.gov/genbank.

3.2. Characterization of the nucleotides of HBV quasi-subgenotype B3

A total of 141 HBV quasi-subgenotype B3 complete genome sequences (34 from this study and 107 from the GenBank) were analyzed. After complete alignment, quasi-subgenotype B3 showed sixteen specific nucleotides differ from B1, B2, B4 and B5 (formerly B6) (Sakamoto et al., 2007), the nucleotides were nt93 T, nt100C, nt355 G, nt843 T, nt861C, nt912C, nt929 T, nt930 G, nt1023 T, nt1041 T, nt2651C, nt2693 T, nt2970C, nt3054A, nt3087A and nt3171 G (positions are given with reference to the EU330989 genome). Except for C93 T (Ala154Val) in the large S gene, T100C (Tyr337His), A355 G (Asn422Asp), A/C929 T (Gln/Ser613Len), A930 G (Gln/Ser613Len), T3054A (Ser250Thr), G3087A (Gly261Ser) and A/T3171 G (Ile/Phe289Val) in the P gene were missense mutation (Supplementary Table 2), the other substitutions were all synonymous mutations. Clinically relevant mutation sites were analyzed. Among 141 sequences of HBV quasi-subgenotype B3, the mutation rate related to vaccine escape and nucleotide analogues-resistant was 0.0%–5.0%, and the mutation rate related to HBV replication was 1.4%–27.0% (Table 2).

3.3. Evolutionary dynamics of HBV quasi-subgenotype B3

Using the relaxed molecular clock, the mean rate of nucleotide substitution estimated for 141 HBV quasi-subgenotyp B3 were 5.556×10^{-4} (95%HPD: 3.642×10^{-4}- 7.748×10^{-4}) substitutions/site/year, 5.612×10^{-4} (95%HPD: 3.581×10^{-4}-7.687×10^{-4}) substitutions/site/year and 5.660×10^{-4} (95%HPD: 3.565×10^{-4}-7.882×10^{-4}) substitutions/site/ year, with the high, middle and low initial estimated substitution rate, respectively. Under these substitution rates, the estimated tMRCA for the HBV quasi-subgenotype B3 whole-gene was around 110 years ago (95% HPD: 1847–1945), and Chinese-origin strains (from Yunnan) could be the parental strains (Table 3). Around 70 years ago (95% HPD: 1916–1964), Chinese strains were introduced to Indonesia on one occasion, then Indonesia strains were introduced to Malaysia and Philippine (Fig. 1). Around 42 years ago (95% HPD: 1932–1992), Chinese strains were introduced to Laos on one occasion. Skyline plot analysis showed an increase in the effective number of HBV quasi-subgenotype B3 infection from the early of 1940s to 1990s and stabled from then on (Fig. 2).

Table 2
Analysis of Clinically relevant mutation sites of hepatitis B virus quasi-subgenotype B3.

ORF	Clinical implication	Mutation	Wild type	Mutant type
S	Vaccine-escape	G145R	140(99.3%)	1(0.7%)
		T116N	141(100%)	0(0.0%)
		P120S	139(98.6%)	2(1.4%)
		I/T126A/N/I/S	141(100%)	0(0.0%)
		Q129H	136(96.5%)	5(3.5%)
		M133 L/I/T	134(95.0%)	7(5.0%)
		K141E	141(100%)	0(0.0%)
		P142S	141(100%)	0(0.0%)
		D144E	140(99.3%)	1(0.7%)
P	LAM-resistance	M204V	140(99.3%)	1(0.7%)
		L80 V/I	141(100%)	0(0.0%)
		I169T	139(98.6%)	2(1.4%)
		V173L	141(100%)	0(0.0%)
		L180W	140(99.3%)	1(0.7%)
		T184S/G	141(100%)	0(0.0%)
		S202T	140(99.3%)	1(0.7%)
		Q215H	140(99.3%)	1(0.7%)
		A181 T/V	141(100%)	0(0.0%)
	ADF-resistance	N236T	141(100%)	0(0.0%)
		A181 V/T	141(100%)	0(0.0%)
		I233V	141(100%)	0(0.0%)
	ETV–resistance	L180W	140(99.3%)	1(0.7%)
		T184 G/S	141(100%)	0(0.0%)
		S202T	140(100%)	1(0.0%)
		M204V	140(99.3%)	1(0.7%)
	TDF–resistance	P177G	141(100%)	0(0.0%)
		A194 T/S	135(95.7%)	6(4.3%)
		F249A	141(100%)	0(0.0%)
PreC/C	HBeAg-negative/Virus replication enhancement /Disease progression/HCC	G1896A	103(73.0%)	38(27.0%)
		G1899A	121(85.8%)	20(14.2%)
		T1753C/A/G	122(86.5%)	19(13.5%)
		A1762T	114(80.9%)	27(19.1%)
		G1764A	118(83.7%)	23(16.3%)
		C1766T	137(97.2%)	4(2.8%)
		T1768A/C	139(98.6%)	2(1.4%)
X	HCC	L130M	113(80.1%)	28(19.9%)
		V131I	116(82.3%)	25(17.7%)

LAM: Lamivudine; ADF: Adefovir; ETV: Entecavir; TDF: Tenofovir; HCC: Hepatocelluar carcinoma.

Table 3
Estimated tMRCA for hepatitis B virusquasi-subgenotype B3 sequences.

Data set	tMRCA(95%HPD)years		
	7.72×10^{-4}	1.0×10^{-5}	2.210^{-6}
global quasi B3(n = 141)	109(67–156)	112(66–159)	111(67–165)
CN-YN(n = 46)	109(67–156)	112(66–159)	111(67–165)
ID(n = 56)	70(48–95)	70(48–96)	69(47–95)
PH(n = 7)	70(48–95)	70(48–96)	69(47–95)
MY(n = 30)	70(48–95)	70(48–96)	69(47–95)
LA(n = 2)	38(19–59)	44(20–80)	45(19–79)

tMRCAs were calculated using three different substitution rates (7.72×10^{-4} (Zhou and Holmes, 2007),1.0×10^{-5} (Torres et al., 2011) and 2.2×10^{-6} (Paraskevis et al., 2013) substitutions per site per year).

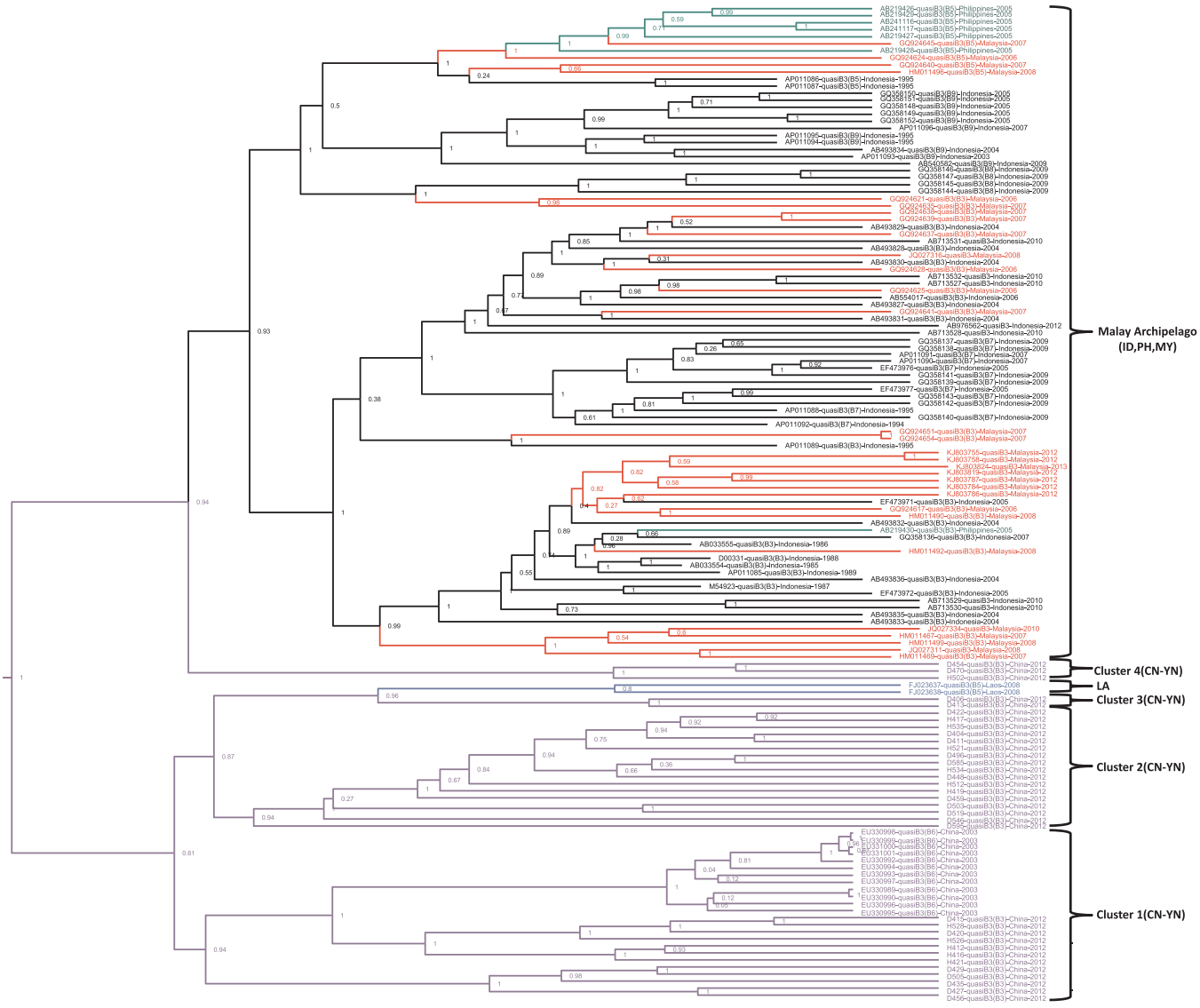

Fig. 1. The Maximum clade credibility (MCC) tree was estimated by a Bayesian analysis of 141 whole genome sequences of hepatitis B virus quasi-subgenotype B3 strains. The posterior probabilities of the key nodes are shown the respective nodes.The HBV quasi-subgenotype B3 sequences obtained from China Yunnan (CN-YN, purple), Indonesia(ID, **black)**, Philippine(PH, green), Malaysia(MY, red)and Laos (LA, blue).

4. Discussion

The heterogeneity of HBV might correlate with HBV replication and HBV seromarkers expression, HBV immune escape, the clinic outcome, the reaction of anti-viral treatment and prognosis. A mean genetic distance of 3.1% within subgenotypes of B3, B5, B7, B8, B9, B6 from China (our previous report) and sequences from this study indicates that all these strains/lineages belongs to a single subgenotype, which concerted with the previous reports and were reclassified as HBV quasi-subgenotype B3 (Shi et al., 2013; Kramvis, 2014). To our knowledge, no structure characteristic of HBV quasi-subgenotype B3 were report until now. In this study, quasi-subgenotype B3 showed sixteen specific nucleotides differ from B1, B2, B4 and B5 (formerly B6). Among them, eight mutations as specific sites of HBV quasi-subgenotype B3, including two at preS1 gene region and six at P gene region resulted in amino acids variation, although their function to biological characteristics of HBV still unknown. Simultaneously, few vaccine escape and nucleotide analogues-resistant variation were observed in HBV quasi-subgenotype B3, which means current vaccine and nucleotide analogues might have effective prevention and treatment for HBV quasi-subgenotype B3.Variations in regions related to HBV genome replication were also observed in HBV quasi-subgenotype B3 during its nature history, whether it could enhance the HBV replication need to be further studied.

Bayesian analysis based on population genetics of current time and nucleotide mutation rates has been widely used in virology evolutionary studies. Using BEAST software, the HBV original time and evolutionary rate were estimated in many studies. For example, HBV C genotype was verified with twice evolutionary rate of genotype B, and origin of genotype B was approximately 60 years earlier than genotype C (Wang et al., 2014). The original native HBV/F2a of Brazil was deduced a north-to-south flow from Venezuela to Brazil (Mello et al., 2013). Subgentype HBV/A1 outside Africa and HBV/E (Afro-Colombian community and remote communities in rural Nigeria) could be traced to Slave trade (Alvarado Mora et al., 2010; Forbi et al., 2010). However, to our knowledge, no Bayesian analysis was done on HBV quasi-subgenotype B3.To estimate the original time and evolutionary rate, we estimated the tMRCA of all collected HBV quasi-subgenotype B3 sequences (141) and reconstructed MCC trees by BEAST software. Our results showed that the estimated evolutionary rate of HBV quasi-subgenotype B3 was 5.556–5.660 $\times 10^{-4}$ substitutes/site/year, and the analyses converge in almost the same evolutionary rate of quasi-B3 no

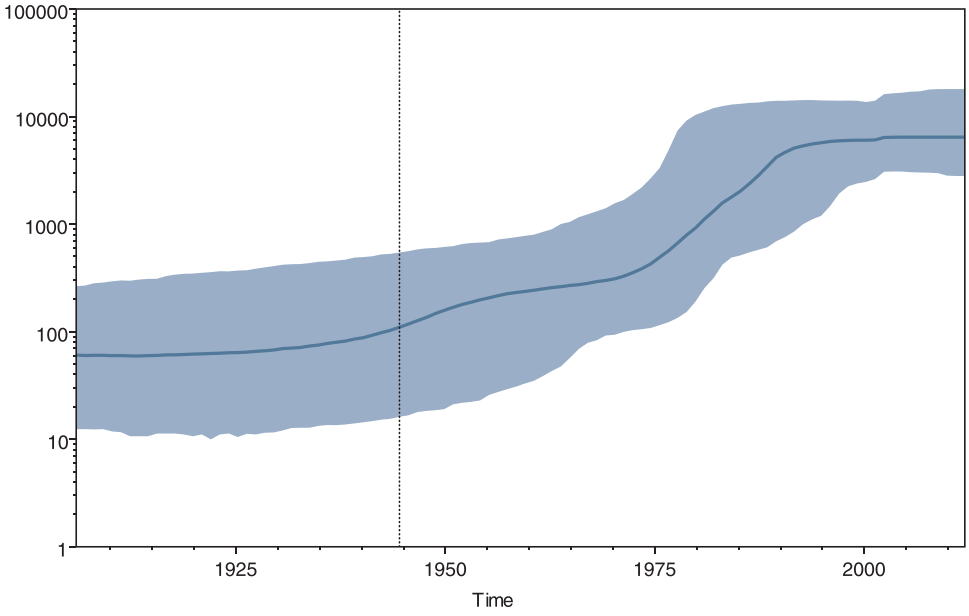

Fig. 2. Bayesian skyline plot showing the epidemic history estimated from the 141 hepatitis B virus quasi-subgenotype B3 dataset. The middle line is the median estimate of effective population size **(log10)**, and the blue area shows the median posterior density intervals for this estimate. The youngest time is the time of collected sequences in this report (2012).

matters what initial values were setting. The evolutionary rate obtained is expected for HBV due to the data included (a single subgenotype) and temporal calibration employed, and is in line with previous estimation for HBV short-term evolutionary rates.In addition, these high rates were associated with tMRCA estimates in this study, ranging from 66-165ybp. However, whether the same evolutionary rate of quasi-B3 was reliable due to the temporal signal, or as an infectious virus and distribution in specific geographic locations, the confounding of temporal and genetic structure was likely to be common still need to be further studied (Firth et al., 2010; Duchene et al., 2015; Murray et al., 2016). MCC tree showed that the quasi-subgenotype B3 was distributed in China (Yunnan) and Southeast Asia, and the median estimated timing of tree's root suggested that quasi-subgenotype B3 was originated in round 1900, Yunnan strains was the earliest to diverge from ingroup. Indonesia, Philippine and Malaysia were date back to earlier 1940s, and Laos was date back to 1970s. Unlike Yunnan strains was one directional transmission as show in MCC tree, Indonesia, Phillippines and Malaysia strains were admixture transmission (e.g. Indonesia strains can firstly split to Indonesia, Malaysia or Philippine strains, and the new Malaysia strains can further resplit to new Malaysia, Indonesia and Phillippine strains, and the new Phillippine strains can resplit to new Malaysia and Phillippine strains), and quasi-subgenotype B3 sequences sampled from several indigenous populations in Malay Archipelago besides China, such as the Malay Penisula, Indonesian tribes, Papua Indonesia, and Philippine islands, form different distinct quasi-subgenotypes B3 (formerly B3, B5, Chinese B6 and B7-B9), indicating a highly admixture pattern among Indonesia, Philippine and Malaysia populations, and there was no recent admixture between Yunnan and other three populations. As three main Malay race countries, Malaysia, Indonesia and Philippines (the geographical regions of Proto-Malay) share similar language and tradition. Proto-Malays were more genetically related to Malays and Chinese (Deng et al., 2014). So far, by biological anthropologic (including genetics) or archaeologic evidence, there were different theories about the origin of the Proto-Malay, like the Yunnan theory, the seafarers theory and the Taiwan theory. Situated in southwest China, Yunnan is a multinational border province and is regarded as a hotspot for the study of the origin of Asian populations. Recently, by Affymetrix Gene Chip Mapping Xba 50 K Array, genotype data showed that the Melayu Jawa (MY-JV) together with Indonesian Jawa (ID-JV) have a very close genetic relationship with the Chinese Yunnan groups (CN-JN and CN-WA) (Hatin et al., 2011), which is consistent with the fact that the tribal Proto-Malays are believed to have migrated from Yunnan, China about 4000–6000 years ago. However, the tMRAC of HBV quasi- subgenotype B3 in this study was around 1847–1945 (95%HPD), which was inconsistent with the earlier human migration events in this area. In fact, Chinese did not enter the Malay Peninsula in significant numbers until the mid-nineteenth century (Raybeck, 1980). The European colonial powers (Portuguese, Dutch and British) brought in large groups of Chinese and Indian immigrant labours to the British Malaya (Malay Peninsula and Borneo) and the Netherlands East Indies (which became Indonesia) with the growth of tin and gold mining and associated service industries. Coincidently, full-length HBV isolates from China (Yunnan-Dai and Hani) in this study appeared as the ancestral strains of quasi-subgenotype B3,which may hypothesized that HBV infection in Aborigines was introduced from Yunnan migrants after the European colonization of South-East Asia during the last one or two hundred years ago. Simultaneously, the topologic tree showed that quasi-subgenotype B3 were geographical clustering into at least six branches from 4 Chinese clusters (Chinese cluster 1–4), a Malay Archipelago(B3, B5,B7-B9) and a Laos, we would expect the "Aboriginal" genotype quasi-subgenotype B3 genetic diversity to be nested within the geographic diversity of Malay Archipelago. Thus, our results may provide a clearly proof that geography is an important factor in HBV evolution. Xishuangbanna is in southern Yunnan Province, bordering with Laos, due to activities like border trade contacts and interracial marriage between Chinese and Laotians, it's not surprising that Laos strains display a close phylogenetic relationship with Chinese strains.

Skyline plot analysis showed a steady increase in population size from early of the 1940's to the 1970's, then a spurt from middle of 1970s to early of 1990s then stabilization. Besides China (Yunnan) strains, HBV quasi-subgenotyp B3 sequences currently were reported mainly from Malaysia, Indonesia and Philippine. As these countries involved wars like World War II, the Pacific War, national independence movements and won independence in 1940s, 1950s and 1960s, respectively. And during the 1970s and '90 s, policy like New Economic Policy (Malysia) and "New Order" regime (Indonesia) promoted the urbanization and industrialization, the corresponding is that a rapid growth of overall national prosperity, migration of rural people and neighbour countries to the cities such as Kuala Lumpur was increasing. Thus, expansion of HBV quasi-subgenotype B3 infection and transmission events pattern among Indonesia, Philippine and Malaysia populations showed on the sky plot and MCC tree was consistent with that history.

In summary, 34 full-length sequences of HBV quasi-subgenotype B3

from Yunnan-Dai and Yunnan-Hani groups were identified in this study. Then reported quasi-subgenotype B3 (formerly B3, B5, Chinese B6 and B7-9) were characteristic analyzed. Sixteen nucleotides differ from B1, B2, B4 and B5(formerly B6) specific belong to the quasi-subgenotype B3. HBV quasi-subgenotype B3 epidemic in south east Asia was originated in around one or two hundred years ago and likely to have spread steadily from 1940s. The limitation of this study is that it was difficult to distinguish whether clinically relevant variations were natural evolution results or after using medicine due to limited information from patients.

Acknowledgements

This work was supported by the Health Bureau of Yunnan Province [grant numbers 2017NS225, 2016NS225, 2018NS260]; and part by the Science and Technology Department of Yunnan Province [grant number 2018DG010]. We are grateful to Xishuangbanna Dai Autonomous Prefecture hospital, Xishuangbanna Dai Autonomous Prefecture for the support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2019.197762.

References

Alvarado Mora, M.V., Romano, C.M., Gomes-Gouvea, M.S., Gutierrez, M.F., Carrilho, F.J., Pinho, J.R., 2010. Molecular epidemiology and genetic diversity of hepatitis B virus genotype E in an isolated Afro-Colombian community. J. Gen. Virol. 91 (Pt 2), 501–508.

Boglione, L., Cusato, J., Cariti, G., Di Perri, G., D'Avolio, A., 2014. The E genotype of hepatitis B: clinical and virological characteristics, and response to interferon. J. Infect. 69 (1), 81–87.

Chook, J.B., Ngeow, Y.F., Yap, S.F., Tan, T.C., Mohamed, R., 2011. Combined use of wild-type HBV precore and high serum iron marker as a potential tool for the prediction of cirrhosis in chronic hepatitis B infection. J. Med. Virol. 83 (4), 594–601.

Deng, L., Hoh, B.P., Lu, D., Fu, R., Phipps, M.E., Li, S., Nur-Shafawati, A.R., Hatin, W.I., Ismail, E., Mokhtar, S.S., Jin, L., Zilfalil, B.A., Marshall, C.R., Scherer, S.W., Al-Mulla, F., Xu, S., 2014. The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia. Hum. Genet. 133 (9), 1169–1185.

Duchene, S., Duchene, D., Holmes, E.C., Ho, S.Y., 2015. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. Mol. Biol. Evol. 32 (7), 1895–1906.

Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., Rambaut, A., 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol. Biol. Evol. 27 (9), 2038–2051.

Forbi, J.C., Vaughan, G., Purdy, M.A., Campo, D.S., Xia, G.L., Ganova-Raeva, L.M., Ramachandran, S., Thai, H., Khudyakov, Y.E., 2010. Epidemic history and evolutionary dynamics of hepatitis B virus infection in two remote communities in rural Nigeria. PLoS One 5 (7), e11615.

Ghosh, S., Banerjee, P., Deny, P., Mondal, R.K., Nandi, M., Roychoudhury, A., Das, K., Banerjee, S., Santra, A., Zoulim, F., Chowdhury, A., Datta, S., 2013. New HBV subgenotype D9, a novel D/C recombinant, identified in patients with chronic HBeAg-negative infection in Eastern India. J. Viral Hepat. 20 (3), 209–218.

Hatin, W.I.N.-S.A., Zahri, M.K., Xu, S., Jin, L., Tan, S.G., Rizman-Idid, M., Zilfalil, B.A., 2011. HUGO pan-asian SNP consortium. Population genetic structure of peninsular malaysia malay sub-ethnic groups. PLoS One 6 (4), e18312.

Hundie, G.B., Stalin Raj, V., Gebre Michael, D., Pas, S.D., Koopmans, M.P., Osterhaus, A.D.M.E., Smits, S.L., Haagmans, B.L., 2017. A novel hepatitis B virus subgenotype D10 circulating in Ethiopia. J. Viral Hepat. 24 (2), 163–173.

Ito, K., Yotsuyanagi, H., Yatsuhashi, H., Karino, Y., Takikawa, Y., Saito, T., Arase, Y., Imazeki, F., Kurosaki, M., Umemura, T., Ichida, T., Toyoda, H., Yoneda, M., Mita, E., Yamamoto, K., Michitaka, K., Maeshiro, T., Tanuma, J., Tanaka, Y., Sugiyama, M., Murata, K., Masaki, N., Mizokami, M., Japanese, A.H.B.S.G., 2014. Risk factors for long-term persistence of serum hepatitis B surface antigen following acute hepatitis B virus infection in Japanese adults. Hepatology 59 (1), 89–97.

Kramvis, A., 2014. Genotypes and genetic variability of hepatitis B virus. Intervirology 57 (3-4), 141–150.

Li, X., Liu, Y., Xin, S., Ji, D., You, S., Hu, J., Zhao, J., Wu, J., Liao, H., Zhang, X.X., Xu, D., 2017. Comparison of detection rate and mutational pattern of drug-resistant mutations between a large cohort of genotype B and genotype C hepatitis B virus-infected patients in North China. Microb. Drug Resist. 23 (4), 516–522.

Meldal, B.H., Bon, A.H., Prati, D., Ayob, Y., Allain, J.P., 2011. Diversity of hepatitis B virus infecting Malaysian candidate blood donors is driven by viral and host factors. J. Viral Hepat. 18 (2), 91–101.

Mello, F.C., Araujo, O.C., Lago, B.V., Motta-Castro, A.R.C., Moraes, M.T.B., Gomes, S.A., Bello, G., Araujo, N.M., 2013. Phylogeography and evolutionary history of hepatitis B virus genotype F in Brazil. Virol. J. 10 (1), 236.

Mulyanto, Depamede S.N., Surayah, K., Tjahyono, A.A., Jirintai, Nagashima S., Takahashi, M., Okamoto, H., 2010. Identification and characterization of novel hepatitis B virus subgenotype C10 in Nusa Tenggara, Indonesia. Arch. Virol. 155 (5), 705–715.

Mulyanto, Depamede S.N., Wahyono, A., Jirintai, Nagashima S., Takahashi, M., Okamoto, H., 2011. Analysis of the full-length genomes of novel hepatitis B virus subgenotypes C11 and C12 in Papua, Indonesia. J. Med. Virol. 83 (1), 54–64.

Mulyanto, Depamede S.N., Surayah, K., Tsuda, F., Ichiyama, K., Takahashi, M., Okamoto, H., 2009. A nationwide molecular epidemiological study on hepatitis B virus in Indonesia: identification of two novel subgenotypes, B8 and C7. Arch. Virol. 154 (7), 1047–1059.

Murray, G.G., Wang, F., Harrison, E.M., Paterson, G.K., Mather, A.E., Harris, S.R., Holmes, M.A., Rambaut, A., Welch, J.J., 2016. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol. Evol. 7 (1), 80–89.

Nagasaki, F., Niitsuma, H., Cervantes, J.G., Chiba, M., Hong, S., Ojima, T., Ueno, Y., Bondoc, E., Kobayashi, K., Ishii, M., Shimosegawa, T., 2006. Analysis of the entire nucleotide sequence of hepatitis B virus genotype B in the Philippines reveals a new subgenotype of genotype B. J. Gen. Virol. 87 (Pt. 5), 1175–1180.

Nurainy, N., Muljono, D.H., Sudoyo, H., Marzuki, S., 2008. Genetic study of hepatitis B virus in Indonesia reveals a new subgenotype of genotype B in east Nusa tenggara. Arch. Virol. 153 (6), 1057–1065.

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miyakawa, Y., Mayumi, M., 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J. Gen. Virol. 69 (10), 2575–2583.

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S.Y., Belshaw, R., Allain, J.P., Hatzakis, A., 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57 (3), 908–916.

Pourkarim, M.R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M., Tacke, F., 2014. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J. Gastroenterol. 20 (23), 7152–7168.

Raybeck, D., 1980. Ethnicity and accommodation: malay-chinese relations in Kelantan, Malaysia. Ethnic Groups (2), 241–268.

Rinonce, H.T., Yano, Y., Utsumi, T., Heriyanto, D.S., Anggorowati, N., Widasari, D.I., Lusida, M.I., Soetjipto, Prasanto H., Hotta, H., Hayashi, Y., 2013. Hepatitis B and C virus infection among hemodialysis patients in yogyakarta, Indonesia: prevalence and molecular evidence for nosocomial transmission. J. Med. Virol. 85 (8), 1348–1361.

Sakamoto, T., Tanaka, Y., Orito, E., Co, J., Clavio, J., Sugauchi, F., Ito, K., Ozasa, A., Quino, A., Ueda, R., Sollano, J., Mizokami, M., 2006. Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C among chronic liver disease patients in the Philippines. J. Gen. Virol. 87 (Pt 7), 1873–1882.

Sakamoto, T., Tanaka, Y., Simonetti, J., Osiowy, C., Borresen, M.L., Koch, A., Kurbanov, F., Sugiyama, M., Minuk, G.Y., McMahon, B.J., Joh, T., Mizokami, M., 2007. Classification of hepatitis B virus genotype B into 2 major types based on characterization of a novel subgenotype in Arctic indigenous populations. J. Infect. Dis. 196 (10), 1487–1492.

Shen, T., Gao, J.M., Zou, Y.L., Dong, H., Yan, X.M., 2009. Novel hepatitis B virus subgenotype in the southern Yunnan Province of China. Intervirology 52 (6), 340–346.

Shen, T., Yan, X.M., Liu, H.X., Zhang, B.X., Li, L., Zhang, J.P., Wang, J.L., Xiao, C.J., 2015. Genotype I of hepatitis B virus was found in east Xishuangbanna, China and molecular dynamics of HBV/I. J. Viral Hepat. 22 (1), 37–45.

Shi, W., Zhu, C., Zheng, W., Carr, M.J., Higgins, D.G., Zhang, Z., 2012. Subgenotype reclassification of genotype B hepatitis B virus. BMC Gastroenterol. 12 (1), 116.

Shi, W., Zhang, Z., Ling, C., Zheng, W., Zhu, C., Carr, M.J., Higgins, D.G., 2013. Hepatitis B virus subgenotyping: history, effects of recombination, misclassifications, and corrections. Infect. Genet. Evol. 16, 355–361.

Suppiah, J., Mohd Zain, R., Bahari, N., Haji Nawi, S., Saat, Z., 2015. G1896A precore mutation and association with HBeAg status, genotype and clinical status in patients with chronic hepatitis B. Hepat. Mon. 15 (10), e31490.

Thedja, M.D., Muljono, D.H., Ie, S.I., Sidarta, E., Turyadi, Verhoef J., Marzuki, S., 2015. Genogeography and immune epitope characteristics of hepatitis B virus genotype C reveals two distinct types: asian and papua-pacific. PLoS One 10 (7), e0132533.

Thedja, M.D., Muljono, D.H., Nurainy, N., Sukowati, C.H., Verhoef, J., Marzuki, S., 2011. Ethnogeographical structure of hepatitis B virus genotype distribution in Indonesia and discovery of a new subgenotype, B9. Arch. Virol. 156 (5), 855–868.

Tian, Q., Jia, J., 2016. Hepatitis B virus genotypes: epidemiological and clinical relevance in Asia. Hepatol. Int. 10 (6), 854–860.

Torres, C., Pineiro y Leone, F.G., Pezzano, S.C., Mbayed, V.A., Campos, R.H., 2011. New perspectives on the evolutionary history of hepatitis B virus genotype F. Mol. Phylogenet. Evol. 59 (1), 114–122.

Utsumi, T., Lusida, M.I., Yano, Y., Nugrahaputra, V.E., Amin, M., Juniastuti, Soetjipto, Hayashi, Y., Hotta, H., 2009. Complete genome sequence and phylogenetic relatedness of hepatitis B virus isolates in Papua, Indonesia. J. Clin. Microbiol. 47 (6), 1842–1847.

Utsumi, T., Wahyuni, R.M., Lusida, M.I., Yano, Y., Priambada, N.P., Amin, M., Purwono, P.B., Istimagfiroh, A., Soetjipto, Brule A., Hotta, H., Hayashi, Y., 2015. Full genome characterization and phylogenetic analysis of hepatitis B virus in gibbons and a caretaker in Central Kalimantan, Indonesia. Arch. Virol. 160 (3), 685–692.

Wang, B., Feng, Y., Li, Z., Duan, H., Zhao, T., Zhang, A., Liu, L., Baloch, Z., Xia, X., 2014. Distribution and diversity of hepatitis B virus genotypes in Yunnan, China. J. Med. Virol. 86 (10), 1675–1682.

Zhou, Y., Holmes, E.C., 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. J. Mol. Evol. 65 (2), 197–205.