Dating the origin and dispersal of global hepatitis B virus genotype C in humans

Tianze Li[1,§], Hui Chen[2,§], Liqin Sun[3], Jiaye Liu[4,5,*]

[1] Department of Epidemiology and Health Statistics, College of Public Health, Zhengzhou University, Zhengzhou, Henan, China;
[2] Center for Disease Control and Prevention of Chinese People's Liberation Army, Beijing, China;
[3] National Clinical Research Center for Infectious Diseases, Shenzhen Third People's Hospital, Southern University of Science and Technology, Shenzhen, China;
[4] School of Public Health, Shenzhen University Health Science Center, Shenzhen, China;
[5] Expanded Program Immunization Division, Shandong Provincial Center for Disease Control and Prevention, Jinan, China.

SUMMARY   Hepatitis B virus genotype C (HBV/C) is one of the most prevalent HBV strains worldwide, especially in the Western Pacific and the South-East Asia. However, the origin and evolutionary timescale of HBV/C remains largely unresolved. We analyzed the evolutionary rate and molecular clock phylogeny of 101 full-genome HBV/C sequences sampled globally using a Bayesian Markov Chain Monte Carlo (MCMC) approach. We inferred the spatiotemporal dynamics of the HBV/C worldwide by the Bayesian Stochastic Search Variable Selection (BSSVS). We found that the estimated mean evolution rate of the HBV/C genotype full-genome was 4.32×10^{-5} subs/site/year (95% highest posterior density 3.02×10^{-6} - 8.97×10^{-5}). Phylogeographic reconstruction was able to identify a single location for the origin of the global HBV/C in Australia around A.D. 715. The subgenotype C4 diverged earliest and mainly circulated in Australia, C1 mainly in Southeast Asia, C2 mainly in East Asia and C3 in Remote Oceania. The effective number of HBV infection presented a rapid exponential increase between the 1760s and 1860s followed by a maintained high level until now. Our study, for the first time, provides an estimated timescale for the HBV/C epidemic, and brings new insight to the dispersal of HBV/C in humans globally. Based on the continuous presence of a highly effective viral population, this study provides further evidence of the challenge from a population-based molecular level to eliminate HBV by 2030, and calls for a concerted effort from policy makers, health providers, and society in the globalized world.

Keywords   Hepatitis B virus, HBV genotype C, evolution, phylogenetic, phylogeographic

1. Introduction

Hepatitis B remains a major worldwide public health problem with approximately 257 million individuals infected with hepatitis B virus (HBV) and more than 94 million chronic hepatitis B (*1*). Current estimates place 29% of cirrhosis and 40% of hepatocellular carcinoma (HCC) can potentially be attributed to HBV infections (*2*). The HBV genome is a 3.2 kb partially double-stranded circular DNA composed of four open reading frames encoding for seven proteins. The DNA backbone is characterized by high variability and diversity due to the self-replication strategy of HBV, selection pressure imposed by the host immune system and other exogenous factors. So far, eight genotypes of HBV (A-H) have been identified defined by a diversity of greater than 8% genetic differences in the complete genome sequence (*3*).

Different genotypes have a clear geographical distribution and many risks associated with disease progression resulting in cirrhosis and HCC. HBV genotype C (HBV/C) is the major genotype circulating in Asia and the Western coast of the USA, which account for a large number of the infections worldwide (*4*). In detail, subgenotype C1 is the most common in Southeast Asia and Southern China; C2 is dominant in East Asia (South Korea and Japan) and the northern part of China; C3 persists in Oceania; and C4 is abundant in the Aborigines in Australia. Compared with genotype A and B, chronic infections with genotype C more commonly result in advanced liver diseases with an increased rate of progression to HCC (*5,6*). Additionally, several studies have suggested that disease outcomes are related to some genetic variants. A recent community-based study in

mainland China confirmed that the frequency of HCC-related mutations in HBV/C was significantly higher than that reported in genotype B cases (*7*).

A myriad of evidence has identified HBV as an ancient virus (*8,9*). Investigating the origin of HBV is crucial because it provides a framework for studying the disease burden, and subsequently an understanding of the evolution of HBV pathogenicity with respect to changes in human population size and life expectancy. Especially when considering the risk of disease progression associated with HBV/C strains, it is vital to understand the molecular evolution and epidemiological history of HBV/C genotype. However, to the best of our knowledge, only a few studies explored the origin and dissemination of HBV/C genotype in some countries (Japan) (*10*) or region (east Asia) (*11*). Based on the hypothesis of coincidence between HBV and human migrations, representative sequences from the whole world will be included in the model for analysis of the origin and dissemination of the HBV/C genotype.

Therefore, based on high-resolution phylogenetic and phylodynamic approaches, the study presented here is aimed at investigating the origin and evolution rates of HBV/C genotype, and to reconstruct its spatial and temporal global dynamics, particular attention focused on the study of subgenotypes C1-C4.

2. Materials and Methods

2.1. Sequence querying strategy and selection

HBV/C whole-genome sequences were retrieved from all uploads to the GenBank database (*https://www.ncbi.nlm.nih.gov*) up to the date May 24, 2019. Full-genome sequences were downloaded in "gb" format including the information GenBank accession number, nucleotide sequence, sequence release year, sampling time and area, and sequence length. Full-genome sequences were included with known sampling time and country. The following sequences were excluded: (1) non-human HBV sequences; (2) expression vector sequences; (3) sequences of patients co-infected with HBV/human immunodeficiency virus; (4) sequences with nucleotide length less than 3215 bp; (5) sequences containing illegal characters (*i.e.* characters other than A, T, C and G); (6) recombination sequences. After excluding ineligible sequences, one was chosen to represent the remaining sequences with similarity over 97 percentage, identical isolation country and year.

2.2. Sequence alignment, genotype, genetic distance and recombination

Nucleotide sequences were aligned using BioEdit software v7.0.5.3. Phylogenetic trees were constructed with MEGA v6.0, using the neighbor-joining method after estimation of genetic distance employing the Kimura two-parameter method. A bootstrapping test was performed with 1,000 duplicates, and the transition/transversion rate was set at 2.0. The online tool jumping profile Hidden Markov Model (jpHMM) (*http://jphmm.gobics.de/submission_hbv*), which specializes in detecting recombination events in the HBV genome, was used to analyze the selected sequences and confirm the genotyping results. We also used SimPlot to validated the recombination repeatedly.

2.3. Phylogenetic analysis

Bayesian Markov Chain Monte Carlo (MCMC) method was used to analyze the evolution rate and molecular clock phylogeny of global HBV/C strains with the BEAST software package version 1.7.5. The time of the most recent common ancestor (tMRCA) with 95% highest posterior density (HPD) of global HBV/C was estimated. The calibration point was the year that each strain was isolated. We used the general time reversible nucleotide substitution model with gamma-distributed rates of variable among sites, which were identified as the best fitting model by jModelTest v2.1.7 on the basis of Akaike Information Criterion. Multiple combinations of molecular clock and coalescent models were explored to select the best fitting model. Finally, runs were performed using the constant size, under the strict clock and Bayesian Skyline Plot molecular clock model. Bayesian MCMC analyses were run with a chain of 60 million steps and sampled every 1,000 steps. Convergence of parameters was identified by TRACER v1.5 with the effective sample size exceeding 200. The Maximum Clade Credibility (MCC) tree was calculated with TreeAnnotator with a burn-in period of 6,000 and then visualized in FigTree v1.4.2.

2.4. Phylogeographic analysis

In order to infer the spatiotemporal dynamics of HBV/C worldwide, the Bayesian Stochastic Search Variable Selection (BSSVS) was used to provide evidence for statistically supported diffusion between state variables under BEAST v1.7.5. The results of BSSVS were summarized using SPREAD v1.0.6, a keyhole markup language (KML) file was generated to identify the major routes of geographic diffusion. Bayes factor (BF) test was used to select the most probable diffusion process. To visualize the geographic dispersal of HBV, the KML file was imported to Google™ Earth to produce a graphical animation of the estimated spatiotemporal pathways of global HBV/C.

3. Result

3.1. Characteristics of included sequences

A total of 101 HBV/C full-genome sequences from

23 countries were included in this analysis (Figures S1 and S2, *http://www.ddtjournal.com/action/getSupplementalData.php?ID=96*). The span of origin time of HBV/C sequences was 54 years with the earliest isolated from India in 1963 and the latest from Bangladesh in November 12, 2017. The highest proportion of isolated sequences was observed in 2013 (12.87%), followed by 2012 (9.90%). The included sequences were isolated from Asia (74.26%), Oceania (18.81%), America (3.96%), Europe (1.98%), and Africa (0.99%). In Asia, the highest proportion of isolated sequences was observed in China (16.83%), followed by Malaysia (12.87%); while in Oceania, the highest proportion of isolated sequences was observed in Australia (7.92%), followed by Papua New Guinea (3.96%). The spatial and temporal distribution of included sequences is summarized in Table 1.

3.2. Estimated evolution rates and tMRCA

The mean evolution rate of HBV/C full-genome sequences was evaluated using 101 isolates. After comparing the strict and relaxed clock models by BF test, the strict clock model was the best fit to the data (2lnBF = 466.45). Under the strict clock model, the estimated mean evolution rate of the HBV/C genotype full-genome was 4.32×10^{-5} subs/site/year (95% HPD 3.02×10^{-6} - 8.97×10^{-5}). In addition, we estimated the tMRCA of all the internal nodes of the MCC tree. In general, the estimated mean tMRCA of the tree root was 1,302 years ago (95% HPD 328-5139 years), which corresponds to the origin date of HBV/C genotype back to A.D. 715 (credibility interval between B.C. 3122-A.D. 1689). In particular, HBV-C4 was the first subgenotype diverging from the root in A.D. 1084 (B.C. 1646-A.D. 1794), followed by C3 with an origin date of A.D. 1211 (B.C. 1079, A.D. 1804). While the origin dates of C1 and C2 were more recent: A.D. 1268 (B.C. 913, A.D. 1820) for C1 and A.D. 1243 (B.C. 996, A.D. 1821) for C2, respectively (Table 2).

3.3. Time-scaled phylogeny and phylogeographic analysis

Phylogenetic analysis suggested that the origin of the global HBV/C genotype may likely have been in Australia. Results of the MCC tree (Figure 1) constructed using Tree Annotator revealed that the global HBV/C genotype segregated into two groups in the early stage of divergence. Lineage 1 consisted of HBV subgenotype C4 and circulated in Australia. Lineage 2 continued to diverge and finally segregated into four groups. Group 1 consisted of HBV subgenotype C1, which was mainly endemic in Southeast Asia, such as Malaysia and Thailand. Group 2 consisted of HBV subgenotype C2, which was mainly endemic in East Asia, such as China, Japan, and Korean. Group 3 consisted of HBV

subgenotype C3, which mainly circulated in Oceania including Australia, New Zealand, and Tonga. Group 4 consisted of HBV subgenotype C10 circulating in Indonesia and Papua New Guinea.

The geographic dispersal of the global HBV/C genotype was conducted with the full-genome sequences and is shown in Figure 2, Figure S3 (*http://www.ddtjournal.com/action/getSupplementalData.php?ID=96*) and Figure S4 (*http://www.ddtjournal.com/action/getSupplementalData.php?ID=96*). Phylogeographic reconstruction was able to identify a single location for the origin of the global HBV/C in Australia around 715. We discovered three main dissemination routes. Route 1 the virus crossed the Torres Strait and arrived in Papua New Guina in 992, and then spread to China in 1159, proceeding to Japan in 1620 and to Korean in 1733. Route 2 the virus crossed the Pacific and arrived in Fiji in 1550, and then spread from Fiji to New Zealand in 1749, and to Tonga in 1931. For route 3 the virus arrived in Malaysia in 1421 from China, and developed a new spreading epicenter in Malaysia, which spread to Thailand in 1615, on to Vietnam in 1641, continuing to Bangladesh in 1829, and finally to Taiwan, China in 1904.

3.4. Population dynamic analysis

The strict clock model and Bayesian skyline plot analysis was performed to reconstruct the evolutionary epidemiology of HBV/C based on the full-genome (Figure 3). Between 1642 and the 1730s, the effective number of new HBV infections remained consistent, then increased slowly until the 1760s, a rapid and exponential increase continued until the 1860s, followed by slow growth until the 1960s and maintained a plateau into the 1980s, finally decreasing gradually since 1992.

4. Discussion

The transmission of hepatitis B has a very long history and remains a major public health concern worldwide. In this study, we performed a complete and comprehensive analysis into the possible origin, temporal and spatial dynamics of global HBV/C based on full-genome sequences obtained from the GenBank database. The findings revealed that (*1*) the mean evolution rate of HBV/C full-genome was 4.32×10^{-5} subs/site/year (95% HPD 3.02×10^{-6} - 8.97×10^{-5}); (*2*) The global HBV/C may have derived from Australia around A.D. 715 and finally segregated into five lineages corresponding to the five subgenotypes; (*3*) The effective number of HBV infections presented a rapid and exponential increase during 1760s and 1860s, and maintained a high level of transmission up to now.

Due to the lack of proof-reading activity of the reverse transcriptase of HBV, nucleotide misincorporation occurs during genome replication,

Table 1. The spatial and temporal distribution of included global full HBV/C sequences

Year	ARG	AUS	BD	BEL	CHN	CHN-HK	CHN-TW	FJ	INA	IND	JPN	KOR	MAS	MYA	NZL	PAN	PHI	PNG	RSA	TG	THA	USA	VN	Total
1963	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
1984	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
1991	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
1992	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
1993	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
1994	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
1996	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
1997	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2
1998	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	4
1999	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
2000	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
2001	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2
2002	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2
2003	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	2	-	-	-	-	1	-	-	5
2004	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	3
2005	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	4
2006	-	-	-	1	1	-	-	-	1	-	1	1	2	-	-	-	-	-	-	-	-	-	-	7
2007	-	-	-	-	3	-	-	-	2	-	1	-	1	-	-	-	-	-	1	-	-	-	-	6
2008	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	2	-	-	5
2009	-	-	-	-	1	-	-	-	-	-	1	1	3	-	-	-	-	-	-	-	1	-	-	7
2010	-	1	-	-	1	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-	-	5
2011	-	1	-	-	-	-	1	-	1	-	1	-	-	-	-	-	-	-	-	-	1	-	-	5
2012	-	1	-	-	4	1	-	-	-	-	1	-	2	-	-	-	-	-	-	-	-	-	1	10
2013	-	1	-	-	1	1	1	-	1	1	1	1	3	-	-	-	1	-	-	-	-	-	1	13
2014	-	1	-	-	-	-	-	-	-	-	1	-	1	1	-	-	-	-	-	-	-	-	1	5
2015	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
2016	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2
2017	-	-	1	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	3
Total	1	8	2	2	17	2	4	2	4	2	10	6	13	1	3	2	1	4	1	2	10	1	3	101

Number in the table means the counts of sequences sampled in corresponding country and year. - means there are no sequences sampled in corresponding country and year. *Abbreviations*: ARG, Argentina; AUS, Australia; BD, Bangladesh; BEL, Belgium; CHN, China; CHN-HK, Hongkong, China; CHN-TW, Taiwan, China; FJ, Fiji; INA, Indonesia; IND, India; JPN, Japan; KOR, Korea; MAS, Malaysia; MYA, Myanmar; NZL, New Zealand; PAN, Panama; PHI, Philippines; PNG, Papua New Guinea; RSA, South Africa; TG, Tonga; THA, Thailand; USA, United States; VN, Viet Nam.

which leads to an inconsistent substitution rate. This study found that the mean evolution rate of HBV/C full-genome was 4.32×10^{-5} subs/site/year (95% HPD 3.02×10^{-6} - 8.97×10^{-5}), which was consistent with the results reported in an analysis based on the global HBV genotype A sequences (3.0×10^{-5} subs/site/year) (*12*), and higher than the result in a study based on the HBV/C sequences from East Asia (5.6×10^{-4} subs/site/year) (*11*). It is well accepted that the estimated time course of the evolution of HBV mainly depends on the nucleotide substitution rate. Furthermore, the substitution rate is dependent on the calibration approach. In particular, the estimated evolution rate may be faster when recent calibration points are used over a period of a few years (*13*), and in contrast, slower when based on more remote points (such as fossil or primate data) (*8,14,15*). This could also account for the difference in the estimated evolution rates between our own study and the results from the East Asia study. Due to calibration errors, model mis-specifications or mutational saturation, especially the fact that not all of the current circulating mutants will maintain fixed in the population, the evolutionary rates

were presumed to change over time (*16*). Theoretically, it is appropriate to use a short-term evolutionary rate to calculate the time-scale of recent events (such as the intra-genotype evolution) and similarly appropriate to use a long-term substitution rate to study events that are distant in time (such as the origin, co-divergence, and cross-species transmission between human and primates). Based on this principle, we focused our studies more on the interaction of HBV/C evolution and population dynamics using sequences from humans, and their public health consequences especially in the modern history of the world.

Although there have been many disputes on the origin of HBV infections in humans, several phylogenetic analyses provided the probable evidence that HBV is an ancient virus based on sequences from human and primates (originated 33,600-34,100 years ago) (*8,17*). Based on the representative sample of currently available sequences from human infections, we inferred that the timing of global HBV/C in humans was estimated to originate 1,302 years ago. The estimated origin time of HBV/C in our study was much earlier than the results in the study based on isolates from East Asia (*11*), which was consistent with the differences in estimated evolutionary rate. In the study reported here we found that the global HBV/C originated in Australia, the C4 subgenotype branched from the rest of genotypes first and circulated in Australia, and the C3 subgenotype spread to Remote Oceania. The phylogeographic scenarios of C3 and C4 were highly consistent with the results from Paraskevis D, *et al.* (*8*). We also found that the virus expanded from Australia

Table 2. Estimates of the origin and evolutionary rate of global full HBV/C sequences

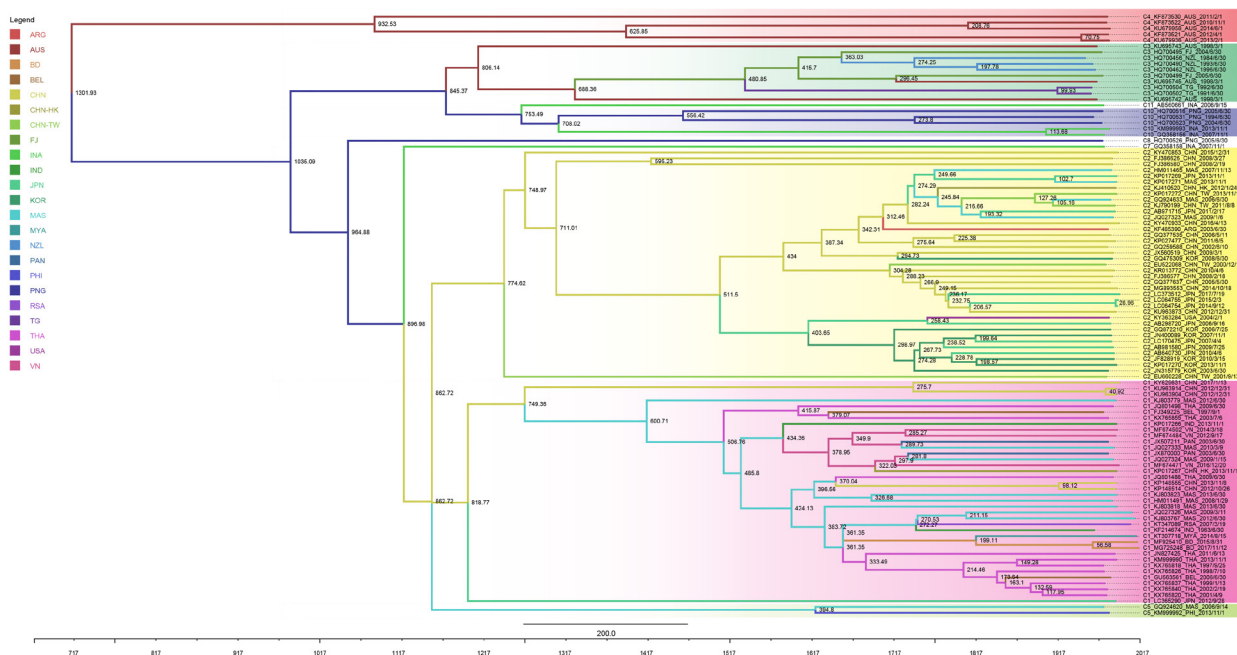| Genotype	Mean value	95% HPD
global C	A.D. 715	B.C. 3122, A.D. 1689
C1	A.D. 1268	B.C. 913, A.D. 1820
C2	A.D. 1243	B.C. 996, A.D. 1821
C3	A.D. 1211	B.C. 1079, A.D. 1804
C4	A.D. 1084	B.C. 1646, A.D. 1794

Figure 1. MCC tree of the complete HBV genotype C sequences throughout the world visualized in FigTree. Dated virus phylogeny displaying subgenotypes within genotype C. The colors of the branches corresponded to their probable geographic location. The intervals of branch reflect the 95% HPD intervals for the branch height. Numbers on the horizontal axis correspond to years before present. Abbreviations of geographic location are shown as described in the note of Table 1.
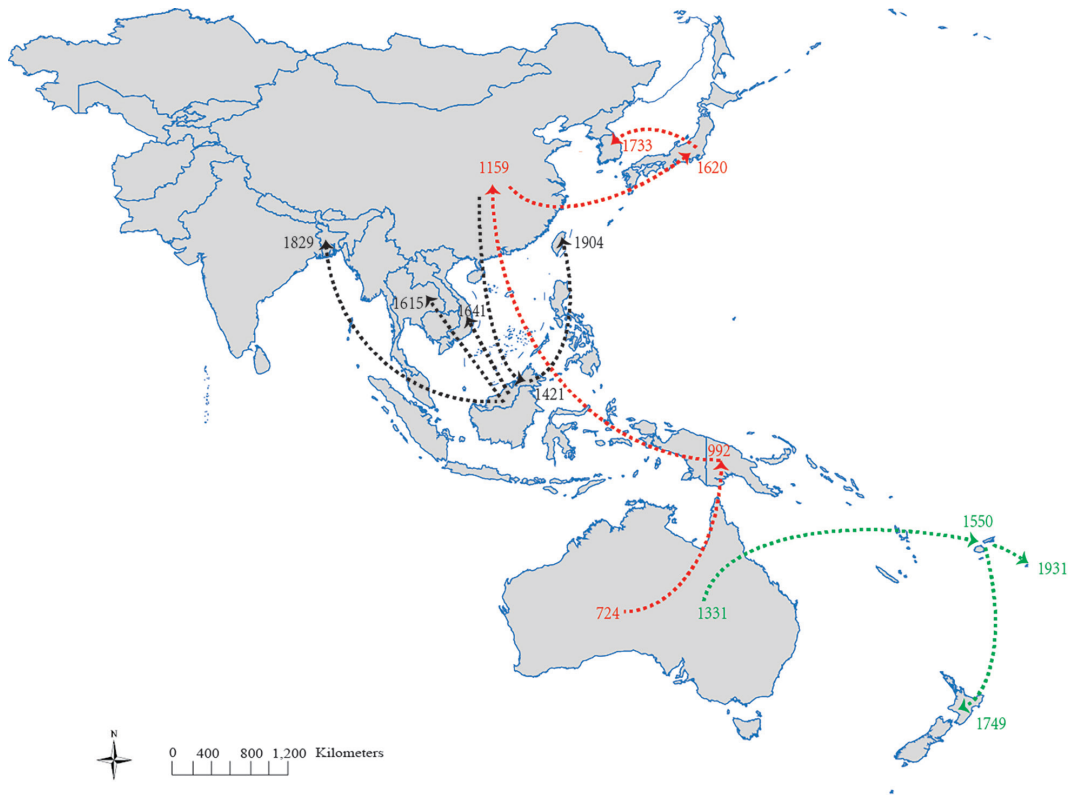
Figure 2. Geographic distribution and inferred dynamics of global HBV genotype C. The map is reconstructed using the ArcGIS (*http://www.esri.com/*), and is identical to the original image created by the SPREAD and Google™ Earth. Arrows indicate the inferred routes of spread of HBV genotype C. The number next to the arrows represent the time when the virus arrived. Main dissemination routes are colored according to their geographic location and time-scale.
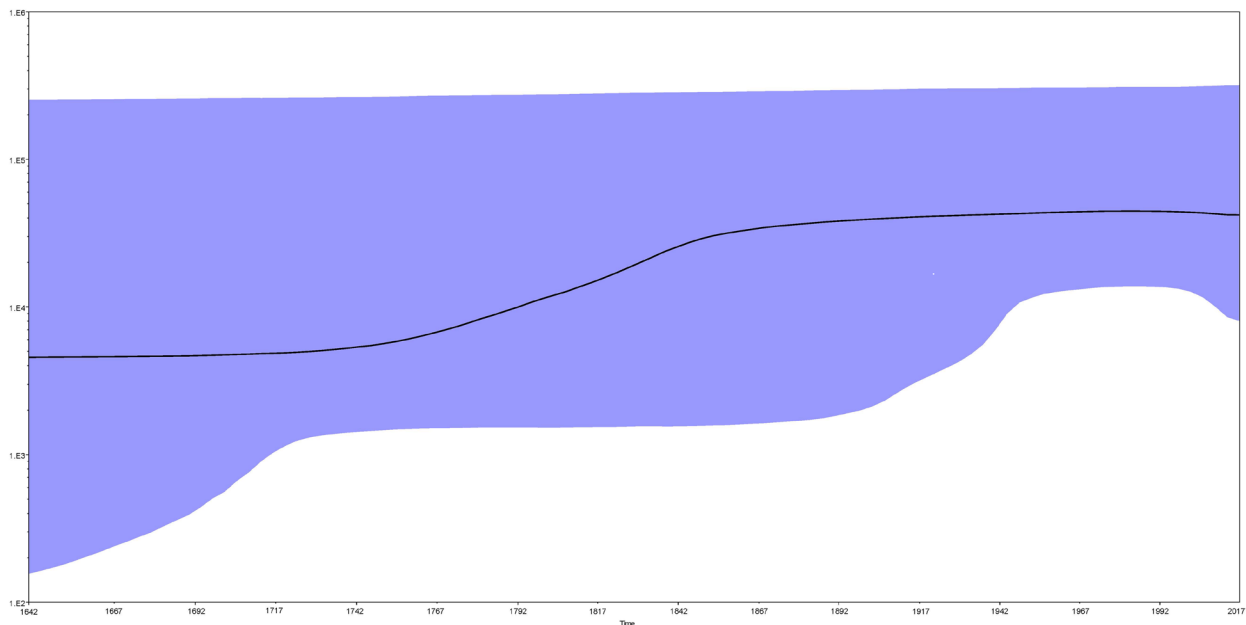

Figure 3. Bayesian skyline plot obtained by analyzing the 101 global HBV/C genotype sequences through time. The solid line stands for the median and shaded areas correspond to the 95% HPD intervals. The effective number of infections is reported on the Y-axis. Time (in years before 2017) is reported on the X-axis.

around A.D. 724, *via* Papua New Guinea, and arrived in China around A.D. 1159, and then spread to Japan in A.D. 1620 and to Korea in A.D. 1733. The detection and molecular characterization of HBV DNA isolated

from a Korean Child naturally mummified in the 16th Century A.D. (*9*), together with frequent trade contacts, diplomatic activities, and wars since Tang Dynasty (7th Century A.D) (*18,19*) supplied possible evidence

to support the credibility of our results. In Southeast Asia, Malaysia became a new disseminating center from which the virus spread to Thailand, Vietnam, and Bangladesh respectively between 17th and 19th century, which could be explained by trade contacts, wars, large scale of population migration, and colonization during the Melaka Sultanate (A.D. 1402-1511) (*20,21*), and the colonial era (A.D. 1511-1914) (*22,23*).

The demographic history of the effective viral population is another important factor that influenced the rate of virus evolution, which comprehensively reflects the scale and dynamics of the host population and the epidemiology/ecological characteristics of infection. In small, stable and isolated populations, the virus may transmit widely and become hyper-endemic. In this setting, the evolutionary rate is usually low, and the dominant route of transmission is vertical. Infected populations frequently presented a high proportion of immunotolerance. On the contrary, HBV when transmitted into a large, highly mobile and susceptible population, will lead to higher evolutionary rates and patients present with a high proportion of immunocompetence. In this study, the Bayesian Skygrid plot indicated that the effective number of HBV infections present a rapid and exponential increase between the 1760s and 1860s, which corresponded to the comprehensive dissemination of HBV in Remote Oceania, Southeast Asia and East Asia as shown by the phylogeographic analysis. Interestingly, the effective viral population began to slowly decline since 1992, possibly due to increasing immunization coverage, the scale-up of antiviral treatment and the prevention of mother to child transmission. However, the continuous high level of the effective viral population remains a large obstacle to end the HBV epidemic.

There are several implications based on this study: (*1*) To the best of our knowledge, this is the first comprehensive analysis of the phylogeographic and phylodynamic spread of the global HBV/C genotype in humans based on a representative sample of current available HBV/C genotype complete sequences. (*2*) The results of this study supplied further evidence that the HBV/C genotype is an ancient virus, chronologically diverged and disseminated companion of the population dynamics, mainly in Oceania, Southeast Asia and East Asia. Especially, the results exhibited an ongoing and concerning condition of higher effective viral population present in the Asian-Pacific region based on molecular evolution. (*3*) In 2017, the World Health Organization (WHO) published a report defining the criteria for hepatitis elimination and outlining a strategy to achieve this goal by 2030 (*24*). However, the latest global hepatitis report revealed that the Western Pacific Region had the highest prevalence of HBV infection and the largest infected population, and the South-East Asia Region has the third highest prevalence of HBV infection. HBV/C is the dominant genotype in the two

regions. Thus, the Western Pacific and the South-East Asia are key regions in achieving the goal of HBV elimination by 2030 due to the heavy burden of HBV infection, which was especially caused by HBV/C genotype. This study also provided molecular evidence of the high effective viral population in these areas. Together with the epidemiologic and molecular data, it calls for a concerted effort from policy makers, health providers, and society in these regions to assist (*25*) with the HBV crisis.

There are some limitations in this study. First, HBV evolution rate, time-scaled phylogeny and phylogeographic analysis are all influenced by the number of included sequences and their isolation location and time. However, not all the countries uploaded HBV/C sequences and we also could not obtain earlier sequences restrained by the development of molecular biological technique. Based on 101 sequences from 23 countries and a timespan of 54 years, the study may underestimate the origin time of HBV/C and provide a probable scenario of molecular evolution. Second, similar to all other studies, we also presumed the evolutionary rate did not change over time, which could influence the precision of our estimation. Thus, further studies are needed to confirm the time-dependency characteristics of the evolutionary rate. Third, recombination revealed the interaction of different virus genotypes and could at times bias evolutionary relationships when constructing phylogenetic trees. Thus, recombination sequences were commonly excluded. However, recombination of different genotypes could also influence the molecular evolution of HBV. Thus, additional innovative analysis methods are also needed to address this issue.

Our study, for the first time, provides an estimated timescale for the HBV/C epidemic, and brings new insight to the dispersal of HBV/C in humans globally. Our study also added additional evidence for the hypothesis of HBV/C divergence and co-expanding with human populations. Based on the continuous condition of high effective viral population, this study further demonstrated a challenge from a population-based molecular level to eliminate HBV by 2030, and calls for a concerted effort from policy makers, health providers, and societies in the globalized world.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in

2016: A modelling study. Lancet Gastroenterol Hepatol. 2018; 3:383-403.

2. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet (London, England). 2018; 392:1736-1788.

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

4. Kramvis A, Kew M, François G. Hepatitis B virus genotypes. Vaccine. 2005; 23:2409-2423.

5. Toan NL, Song le H, Kremsner PG, Duy DN, Binh VQ, Koeberlein B, Kaiser S, Kandolf R, Torresi J, Bock CT. Impact of the hepatitis B virus genotype and genotype mixtures on the course of liver disease in Vietnam. Hepatology. 2006; 43:1375-1384.

6. Lin CL, Kao JH. The clinical implications of hepatitis B virus genotype: Recent advances. J Gastroenterol Hepatol. 2011; 26 Suppl 1:123-130.

7. Yin J, Zhang H, He Y, Xie J, Liu S, Chang W, Tan X, Gu C, Lu W, Wang H, Bi S, Cui F, Liang X, Schaefer S, Cao G. Distribution and hepatocellular carcinoma-related viral properties of hepatitis B virus genotypes in Mainland China: a community-based study. Cancer Epidemiol Biomarkers Prev. 2010; 19:777-786.

8. Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Allain JP, Hatzakis A. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology. 2013; 57:908-916.

9. Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, Kim TH, Kim SB, Grant PR. Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology. 2012; 56:1671-1680.

10. Lin SY, Toyoda H, Kumada T, Liu HF. Molecular evolution and phylodynamics of acute hepatitis B virus in Japan. PloS One. 2016; 11:e0157103.

11. Lin SYC, Magalis BR, Salemi M, Liu HF. Origin and dissemination of hepatitis B virus genotype C in East Asia revealed by phylodynamic analysis and historical correlates. J Viral Hepat. 2019; 26:145-154.

12. Zehender G, Svicher V, Gabanelli E, *et al*. Reliable timescale inference of HBV genotype A origin and phylodynamics. Infect Genet Evol. 2015; 32:361-369.

13. Zehender G, De Maddalena C, Giambelli C, Milazzo L, Schiavini M, Bruno R, Tanzi E, Galli M. Different evolutionary rates and epidemic growth of hepatitis B virus genotypes A and D. Virology. 2008; 380:84-90.

14. Godoy BA, Alvarado-Mora MV, Gomes-Gouvêa MS, Pinho JR, Fagundes N, Jr. Origin of HBV and its arrival in the Americas--the importance of natural selection on time estimates. Antivir Ther. 2013; 18:505-512.

15. Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. Time-dependent rates of molecular evolution. Mol Ecol. 2011; 20:3087-3101.

16. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. Syst Biol. 2007; 56:515-522.

17. Paraskevis D, Angelis K, Magiorkinis G, Kostaki E, Ho SY, Hatzakis A. Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes. Mol Phylogenet Evol. 2015; 93:44-54.

18. Wei CC. China's cultural routes to korean peninsular and japan and its cultural exchange in the han dynasty and tang dynasty. Jilin Da Xue She Hui Ke Xue Xue Bao. 2008; 1:89-95,159-160. (in Chinese)

19. Atwell WS. Some observations on the "seventeenth-century crisis" in china and japan. J Asian Stud. 1986; 45:223-244.

20. Wilkinson RJ. The Malacca Sultanate. Journal of the Malayan Branch of the Royal Asiatic Society. 1935; 13:22-67.

21. Borschberg P. The Melaka Empire. c.1400-1528. Brill, Leiden, Netherlands, 2019.

22. Andaya, Leonard Y. A history of Malaysia. St Martin's Press. 1982.

23. Hack K, Rettig T. Colonial armies in Southeast Asia. Routledge, London, UK, 2006.

24. World Health Organization. Combating hepatitis B and C to reach elimination by 2030, Advocacy brief. Geneva. *http://apps.who.int/iris/bitstream/10665/206453/1/WHO_HIV_2016.04_eng.pdf; ua=1* (Accessed 23 Oct 2019).

25. World Health Organization. Global hepatitis report, Geneva. 2017.

§These authors contributed equally to this work.
Address correspondence to:
Jiaye Liu, School of Public Health, Shenzhen University Health Science Center, Shenzhen, 518060, China.
E-mail: liujiaye1984@163.com