Received: 22 June 2018 | Accepted: 13 August 2018

DOI: 10.1111/jvh.13006

ORIGINAL ARTICLE

WILEY JOURNAL OF VIRAL HEPATITIS JVH

Origin and dissemination of hepatitis B virus genotype C in East Asia revealed by phylodynamic analysis and historical correlates

Serena Y. C. Lin[1] | Brittany Rife Magalis[2,3] | Marco Salemi[4,5] | Hsin-Fu Liu[6,7,8]

[1]Hepatobiliary Section, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[2]Department of Biology, College of Science and Technology, Temple University, Philadelphia, Pennsylvania

[3]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania

[4]Department of Pathology, Immunology and Laboratory Medicine College of Medicine, University of Florida, Gainesville, Florida

[5]Emerging Pathogens Institute, University of Florida, Gainesville, Florida

[6]Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan

[7]Department of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan

[8]Department of Nursing, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

Correspondence
Hsin-Fu Liu, Department of Medical Research, Mackay Memorial Hospital, Taiwan.
Email: hsinfu@mmh.org.tw

Funding information
Ministry of Science and Technology, Taiwan, Grant/Award Number: NSC 102-2917-I-019-001; Mackay Memorial Hospital, Taiwan, Grant/Award Number: MMH-106-125 and 2017-C01-00152

Summary

Hepatitis B virus disease progression in East Asia is most frequently associated with genotype C (HBV/C). The increasing availability of HBV/C genetic sequences and detailed annotations provides an opportunity to investigate the epidemiological factors underlying its evolutionary history. In this study, the Bayesian phylogeography framework was used to investigate the origins and patterns in spatial dissemination of HBV/C by analyzing East Asian sequences obtained from 1992 to 2010. The most recent common ancestor of HBV/C was traced back to the early 1900s in China, where it eventually diverged into two major lineages during the 1930s-1960s that gave rise to distinct epidemic waves spreading exponentially to other East Asian countries and the USA. Demographic inference of viral effective population size over time indicated similar dynamics for both lineages, characterized by exponential growth since the early 1980s, followed by a significant bottleneck in 2003 and another increase after 2004. Although additional factors cannot be ruled out, we provide evidence to suggest this bottleneck was the result of limited human movement from/to China during the SARS outbreak in 2003. This is the first extensive evolutionary study of HBV/C in East Asia as well as the first to assess more realistic spatial ecological influences between co-circulating infectious diseases.

KEYWORDS
East Asia, HBV genotype C, human mobility, phylogeography, population bottleneck, SARS

1 | INTRODUCTION

Hepatitis B virus is a worldwide pathogen highly endemic in Eastern Asia (especially Taiwan, Japan, Korea and China) and is associated with increased risk of liver cirrhosis (LC) and hepatocellular carcinoma (HCC).[1] According to World Health Organization (WHO) data, at least 240 million people are currently suffering from chronic HBV infection, with more than 780 000 deaths per year.[1] The virus has a partially double-stranded DNA genome of approximately 3200 nucleotides, although virions also contain pre-genomic RNA. At the molecular level, HBV is characterized by significant genetic heterogeneity comprising ten genotypes, A to J, with average inter-genotype nucleotide diversity of 8%.[2] Genotype C is further divided into four sub-genotypes with ~4% nucleotide diversity between them.[3]

Previous molecular epidemiology studies based on phylogenetic analysis showed a specific regional distribution pattern of

the different genotypes, except for genotypes D and G, which are scattered worldwide.[4] Genotype A is prevalent in Africa, North America and Europe; Genotypes B and C are the major genotypes circulating in Asia[4] and, even in the USA, are the most common among Asian patients[5]; genotype E is prevalent in Africa; F/H in Central and South America; I in Taiwan; and J in Japan.[4,6,7] HBV genotype C (HBV/C), in particular, is the most prevalent genotype in almost every East Asian country, and it also accounts for a large number of infections in the USA (prevalence of 41% and 23% along the Western and Eastern coasts, respectively[5]). In Korea, genotype C constitutes almost 100% of the infections,[8] approximately 50% in Hong Kong,[9] and 85% in Japan.[10] In Mainland China, genotype C is predominant in the Northern part of China (Beijing, Xingjiang and Gansu), while genotype B is prevalent in the Central and Eastern part of China (Hunan and Fujian), with an overall prevalence of 41% for genotype B and 53% for genotype C.[11] Similarly in Taiwan, where HBV genotype B is the most prevalent (68%), genotype C still accounts for almost one-third of the infections (32%).[12] Several sub-genotypes with specific molecular epidemiology patterns have been described within HBV/C. Sub-genotype C1 (or Cs: s for Southeast Asia) is dominant in Southeast Asia and Southern China; C2 (or Ce: e for East Asia) is found in East Asia (South Korea and Japan) and northern part of China[13]; C3 in Oceania[3]; and C4 in the Aborigines from Australia.[14]

Chronic infection with HBV/C has been associated with significantly higher risk than with other genotypes for progression to LC and HCC.[15,16] In Taiwan, in particular, chronic hepatitis and HCC are still the 9th ranking cause of death.[16] Indeed, several studies have suggested that disease outcomes are related to specific genetic variants.[17-19] Due to the disease activity and the risk of HCC development with which HBV/C is associated, it is important to investigate the molecular evolution and demographic history of this genotype in highly endemic countries notably the high prevalence of genotype C in USA and the prevalence (14.8%) of chronic HBV infections among Asian immigrants who likely acquired infections in their country of origin.[20] We consider this as an Asian-related infection network, and therefore the HBV/C sequences from USA will be included in the study. Previous studies based on Bayesian coalescent analysis estimated the HBV evolutionary rate to be approximately 10^{-4}-10^{-5} nucleotide substitutions/site/year[21,22] and traced back the time of the most recent common ancestor (tMRCA) of the currently circulating human genotypes to ~1500 years ago, which in turn separated from the avian HBV lineage ~6000 years ago.[21] However, specific genotypes, or lineages within genotypes responsible for current outbreaks, may have a more recent origin, and their successful spread could be the result of specific historical or geopolitical correlates during the past decades potentially related to an unprecedented increase in human mobility. Therefore, our main objective was to infer the origin and epidemic history of HBV/C in East Asia and investigate ecological factors affecting dissemination and epidemic outbreaks of the virus.

2 | MATERIALS AND METHODS

2.1 | HBV data set selection

We mined the GenBank database (https://www.ncbi.nlm.nih.gov) to compile a comprehensive data set of all currently available HBV/C sequences. The gold standard method for HBV genotyping is whole-genome sequencing followed by phylogenetic analysis.[23] Therefore, to infer a reliable demographic history of HBV/C, we focused on full-genome sequences with known sampling time and country of origin (both are necessary for calibration of molecular clock and Bayesian inference of spatiotemporal dynamics [see section 2.3]). Sequences lacking isolation time were complemented by the kind email reply from the submitted authors. Sequences with only publication year instead of isolation time were excluded from the data set. To make the most abundant time information to the data set, we kept at least 2 sequences for each isolation year as possible. If insufficient qualified sequence could be obtained from Genbank in a certain year, the sequence number was less.

Sequences included in the final data set were required to satisfy the following criteria: nonrecombinant sequences with no uncertainty concerning genotype assignment; sequences isolated only from human serum or plasma (sequences from liver tumour were excluded to avoid the potential confounding factor of tissue-specific convergent evolution of sequences sampled from different patients); sequences not epidemiologically linked (ie, not linked through a direct transmission chain); and when multiple sequences from the same subjects were available, only one sequence was randomly selected. Genotyping classification was confirmed by phylogenetic analysis using Neighbor Joining (NJ) tree reconstruction, with the GTR nucleotide substitution model, gamma-distributed rate heterogeneity among sites (GTR + G), and 1000 bootstrap replicates, from an alignment including the full-genome sequences obtained from GenBank and well-established genotype (A to J) reference strains (accession numbers shown in Table S1). Calculations were carried out within MEGA6 software.[24] Data collection times of HBV sequences included in the full-genome data set (n = 429) spanned from 1992 to 2010. To ensure that a specific country was not falsely over-represented in the alignment, a sampling ratio was calculated using the proportion of genotype C in chronic HBV cases (China 53%, Korea 98%, Japan 85%, Taiwan 32%, USA 41%) normalized by HBV prevalence in the general population of each country (China 10%,[25] Korea 5.9%,[26] Japan 4%,[27] Taiwan 15%,[28] USA 15%,[20,29] which resulted in a sampling ratio China:Korea:Japan:Taiwan:USA of 1.6:1.7:1:1.4:1.8. Sequences were, then, randomly selected from each country according to this ratio to generate an alignment including a final alignment of 120 strains representative of the virus prevalence in each country spanning from 1992 to 2010 (see Table S2) to infer maximum likelihood (ML) trees, NJ trees and Bayesian coalescent inference.

HBV/C sequences were aligned with the MUSCLE algorithm implemented in the MEGA6 software. Fasta files were downloaded from the link. (https://drive.google.com/

open?id=1Th9vniFFZzo7CLyZ9IYHqoyBfp2mpjQq). Genotype B sequences were used as outgroups (EU919173-EU919176) for rooting the ML and NJ trees. Outgroups were excluded in the data set of the Bayesian inference, as rooting was inferred by calibrating a molecular clock (see below). The alignments are available from the authors upon request.

2.2 | NJ and ML phylogeny and gene flow analysis

The best-fitting nucleotide substitution (GTR + G) model was selected using a hierarchical likelihood ratio test within PAUP* v4.0.[30] NJ and ML trees were then inferred according to the best-fitting model using MEGA6 and PhyML 3.0 (http://www.atgc-montpellier.fr/phyml/), respectively. Statistical supports for internal branches in the phylogeny were assessed by bootstrapping (1000 replicates) and the approximate likelihood ratio test (aLRT). Genotype B sequences were used for outgroup-rooting in the NJ and ML trees. Phylogenies were then annotated with FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

The existence of distinct subpopulations of HBV/C in different countries was analysed by the Slatkin and Maddison[31] test and percentage of viral migrations from/to different countries inferred by maximum parsimonious reconstruction of ancestral states using the MacClade v4.08 program.

2.3 | Molecular clock signal test and Bayesian coalescent inference

To assess the molecular clock signal carried in the temporally sampled viral sequences, a cross-platform software, TempEst (formerly known as Path-O-Gen; http://tree.bio.ed.ac.uk/software/tempest/), is used to explore the association between genetic divergence through time and the sampling dates.

Time-scaled phylogenetic trees, evolutionary rates and demographic histories of HBV/C strains were evaluated using the Bayesian coalescent framework implemented in BEAST v1.8.2 (http://beast.community/index.html), which uses a Markov Chain Monte Carlo (MCMC) sampling method to obtain posterior distributions of tree topologies and parameter estimates. Bifurcating nodes with posterior probability greater than 0.95 were considered statistically well supported. Six different evolutionary models were tested: strict vs relaxed molecular clock, each one with a constant size, exponential growth or Bayesian Skygrid (nonparametric) demographic prior.[32] Depending on the model, MCMCs were run for 500 million to 1.5 billion generations (sampling every 0.01% of the run) until the effective sampling size of each parameter estimate (after burn-in of 10%-25%, depending on the model) was >200 to ensure proper mixing of the Markov chain. For each run, the marginal likelihood was estimated via path sampling (PS) and stepping stone (SS) methods[33] and the resulting Bayes Factors (BF) (ratio of marginal likelihoods) used to select the best-fitting clock/demographic model.[34] In practice, following the original work of Kass and Raftery,[35] any two models can be compared to evaluate the strength of evidence against the null hypothesis (H_0) defined in the following way: $2lnBF < 2$ indicates no evidence against H_0; 2-6, weak evidence; 6-10: strong evidence, and >10 very strong evidence. Maximum clade credibility (MCC) trees were obtained with TreeAnnotator v1.8.2 in the BEAST v1.8.2 software package and edited for display purposes in FigTree.

2.4 | Bayesian phylogeography

In ML and NJ trees, the putative location of each ancestral lineage (internal branch) was inferred by assigning a discrete character to each sequence corresponding to the country of origin and reconstructing ancestral states by maximum parsimony. A more in-depth phylogeographic analysis, incorporating both spatial and temporal information, was also performed with BEAST[36] using a discrete trait, symmetric substitution model with Bayesian stochastic search variable selection (BSSVS). The MCC tree was converted to a keyhole markup language file (KML file) using SPREAD[37] software and projected onto a geographical map using Google Earth (available online: http://www.google.com/earth) to produce a graphical animation of the estimated spatiotemporal patterns of HBV/C evolution. Longitude and latitude of each centre of the cities or countries were marked orderly in a text-delimited file for SPREAD.

2.5 | Human mobility data in East Asia

The entry numbers at customs arrival point of Eastern Asian countries, as well as departure trends, were collected from a number of national databases (see below) to evaluate potential correlations between viral demographic history and human mobility during the past two to three decades (depending on available data for each country). Human mobility data in Taiwan, collected from 1992 to 2015, were downloaded from the Tourism Bureau, Ministry of Transportation and Communications, Taiwan (http://admin.taiwan.net.tw/statistics/year.aspx?no=134.); data in China, collected from 1981 to 2015, were obtained from China National Tourism Administration (http://www.cnta.gov.cn); data in Japan, collected from 1975 to 2015, were obtained from Japan Tourism Agency (http://www.mlit.go.jp/kankocho/zh-tw/siryou/toukei/in_out.html.); data in South Korea, collected from 1995 to 2015, were obtained from Korea Tourism Organization (http://kto.visitkorea.or.kr/eng/tourismStatics/keyFacts/KoreaMonthlyStatistics/eng/inout/inout.kto). Statistical analysis for potential correlations between human mobility and viral population demographic history was assessed by bivariate Pearson correlation using IBM SPSS statistics v.21 software package.

3 | RESULTS

3.1 | HBV/C phylogeny and viral gene flow

The NJ tree inferred from genotype C full-genome sequences was rooted with genotype B strains (see Figure 1). Chinese sequences of HBV/C are closest to the root, followed by lineages belonging to Taiwan, Japan and the west coast of the USA. The Japanese strains

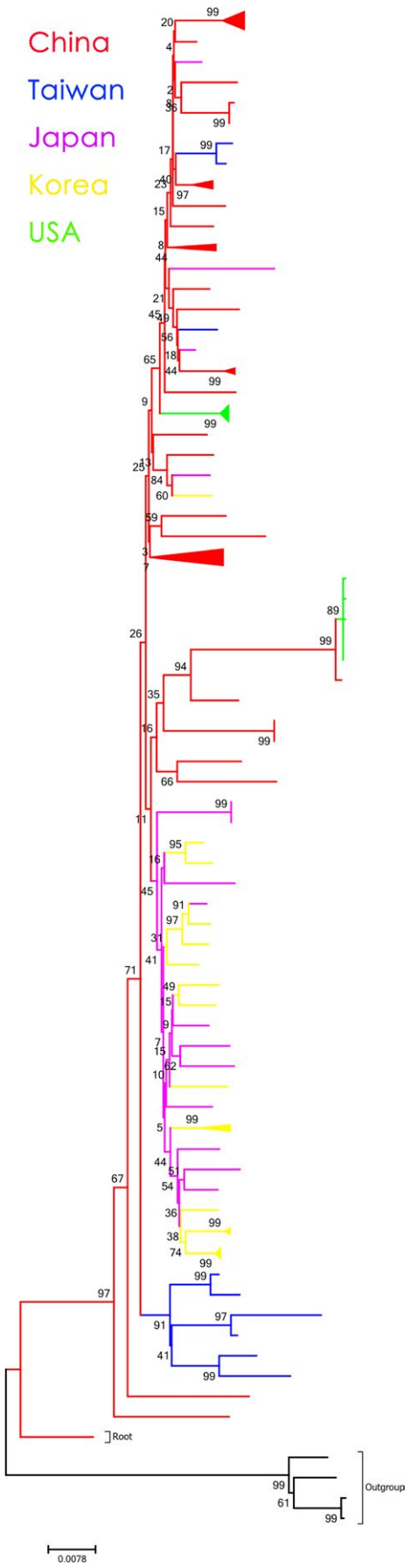

FIGURE 1 NJ tree reconstructed of HBV/C full-genome sequences from East Asia countries and USA. The bootstrap values on branches provide statistical support of lineage clustering. Phylogeographic distribution and direction were determined using outgroup-rooting of genotype B sequences. Countries are specified according to colour. Clades of lineages belonging to the same geographic location were compressed into cartoon triangles

gave further rise to lineages belonging to Korea. The results were confirmed by the ML tree (Figure S1), which exhibits a similar topology. Viral gene flow analysis based on maximum parsimony reconstruction of ancestral location characters (ie, most parsimonious geographic origin of ancestral sequences) shows that the majority of spatial transition events observed in the ML and NJ trees (72.5%-58.8%) occurred from China to other East Asian countries (Figure 2), while the remaining transitions (nearer the tips) occurred between Korea and Japan. Overall, the results implicate China as the initial epicentre of the epidemic from where different lineages progressively spread to other East Asian countries.

3.2 | HBV/C Bayesian phylogeography

In order to investigate the timeframe of HBV/C spatial dispersion patterns in East Asia within the Bayesian coalescent framework, we selected from all available full-genome sequences a random sub-sample according to the relative ratio of HBV/C prevalence in each country (Table S1). A root-to-tip regression plot was inferred from the ML tree to test for the presence of temporal signal required for accurate reconstruction of node ages within the Bayesian phylogeny. The regression slope was positive, with a correlation

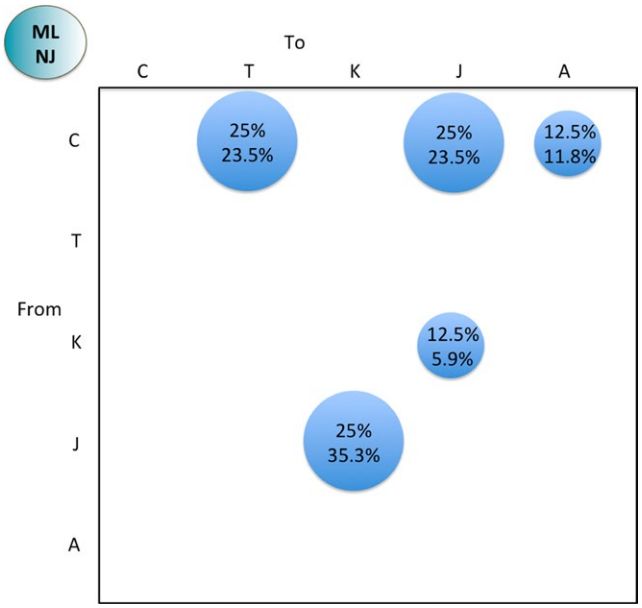

FIGURE 2 HBV/C migration among East Asian countries. Each circle displays the percentage of observed spatial transitions (migrations) inferred from the maximum likelihood (upper value) and Neighbor Joining (lower value) phylogenies. Letters C, T, K, J and A represent China, Taiwan, Korea, Japan and USA, respectively

TABLE 1 Bayes Factors comparison of coalescent models for the HBV/C data set

Data set	N[a]	Model[b]	PS[c]	SS[d]
Data set	120	R_Const	−24107.8	−24110.8
		R_Skygrid	**−22592.0**	**−22593.7**
		S_Const	−24219.9	−24221.3
		S_Skygrid	−22696.6	−22698.0
Lineage 2	136	R_Const	−25333.9	−25335.2
		R_Exp	−25303.6	−25304.9
		R_Skygrid	**−25264.0**	**−25267.6**
Lineage 3	99	R_Const	−15343.3	−15344.7
		R_Exp	−15340.0	−15340.4
		R_Skygrid	**−15334.5**	**−15336.0**

The best-fitting model for each data set is highlighted in bold. Highlighted models all have Bayes Factors >10 compared to any alternative model, indicating strong statistical support.

[a]Number of sequences included in the data set.

[b]Bayesian coalescent model (molecular clock model_demographic prior): S = strict clock, R = relaxed clock, Cost = constant population size, Exp = exponential population growth, Skygrid = nonparametric Skygrid population prior.

[c]Marginal likelihood of the model estimated via path sampling method.

[d]Marginal likelihood of the model estimated via stepping stone sampling method.

coefficient being 0.151, which indicates sufficient temporal signal and suggests that a relaxed molecular clock is likely more appropriate than a strict clock model to explain HBV/C evolution. Indeed, Bayes Factor comparison of different clock models implemented in BEAST, using PS/SS methods to estimate the marginal likelihood, indicated that the relaxed molecular clock with gamma distribution and nonparametric Skygrid demographic models best fit the data (Table 1). Evolutionary history reconstruction using these models revealed a median evolutionary rate for the HBV/C full genome of 5.6×10^{-4} subs/site/year (95%HPD: 4.0×10^{-4}-8.0×10^{-4} subs/site/year), in agreement with previous estimates.[21] The overall topology of the Bayesian maximum clade credibility (MCC) tree (Figure 3A) inferred from the data set is almost the same as ML and NJ phylogenies inferred from the full data set (Figure 1 and Figure S1). The majority of the available strains (99.5%) clustered within lineages 2 and 3, with lineage 2 primarily consisting of Japan and Korean isolates (60.9% of all available sequences within the clade), and lineage 3 of Chinese/Taiwanese isolates (70.5% of all available sequences within the clade) (Figure 3B). The HBV/C most recent common ancestor (MRCA) emerged in China in the mid-1940s with a 95% high posterior density interval (95% HPD) of 1901-1969. After an initial spread in China, three lineages emerged almost simultaneously around the 1950s (95%HPD 1932-1970, 1942-1970 and 1944-1970, respectively [Figure 4]). Lineage 1 gave rise to two separate introductions from China to Taiwan and from China to USA. Lineage 2 exhibited a transition from China to Japan, from where it disseminated in turn to Korea through several separate introductions. Lineage 3 appeared to circulate within China, from where it was further disseminated to Taiwan, through at least three separate introductions, and then to USA. No introduction from Japan into Taiwan was observed. Significantly supported viral migration routes (BF > 3) over 1992-2010, representing the spatial transitions in the MCC tree, are shown in Figure 4.

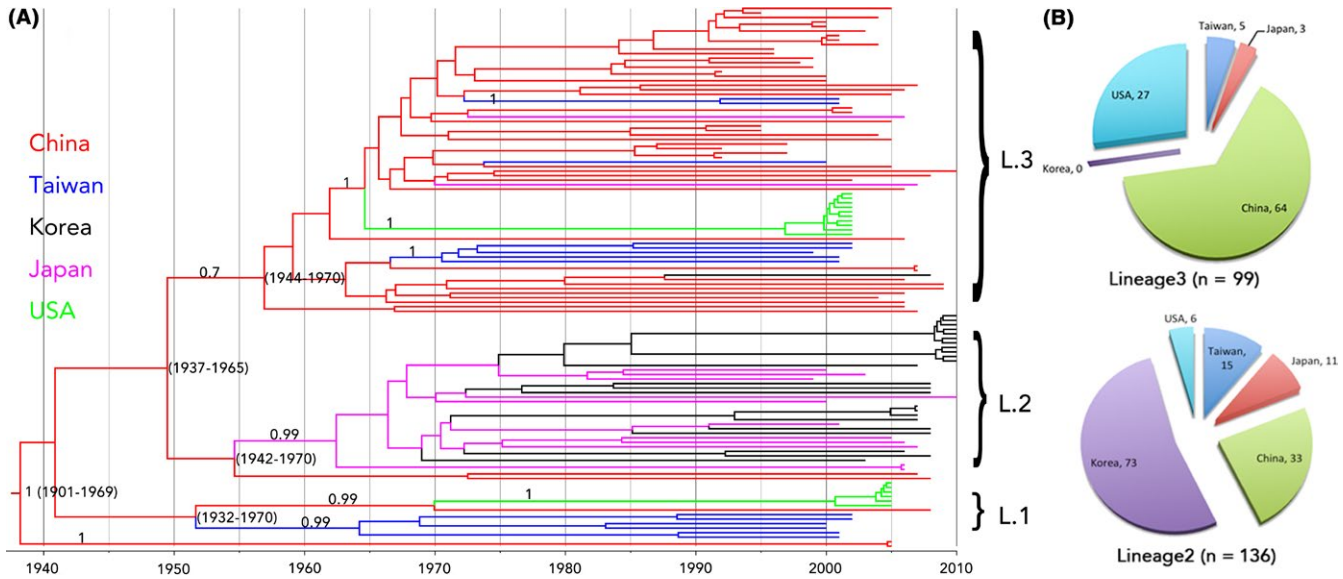

FIGURE 3 Bayesian MCC tree of HBV/C full-genome sequences from East Asian countries. (A) The time-scaled MCC tree was inferred using the Bayesian coalescent framework. Numbers marked on branches represent posterior probabilities. Branches are coloured according to country of origin of each lineage (legend to the left). The tree was scaled in time by calibrating a relaxed molecular clock (lognormal distribution?). The 95% high posterior density intervals for the time of origin of the main ancestral lineages are given in parenthesis to the right of each internal node (ie, inferred ancestor of the lineage). Three major lineages (L.1, L.2 and L.3) originating from China are indicated. (B) Sequence distribution of strains sorted from the MCC tree Lineages 2 and 3

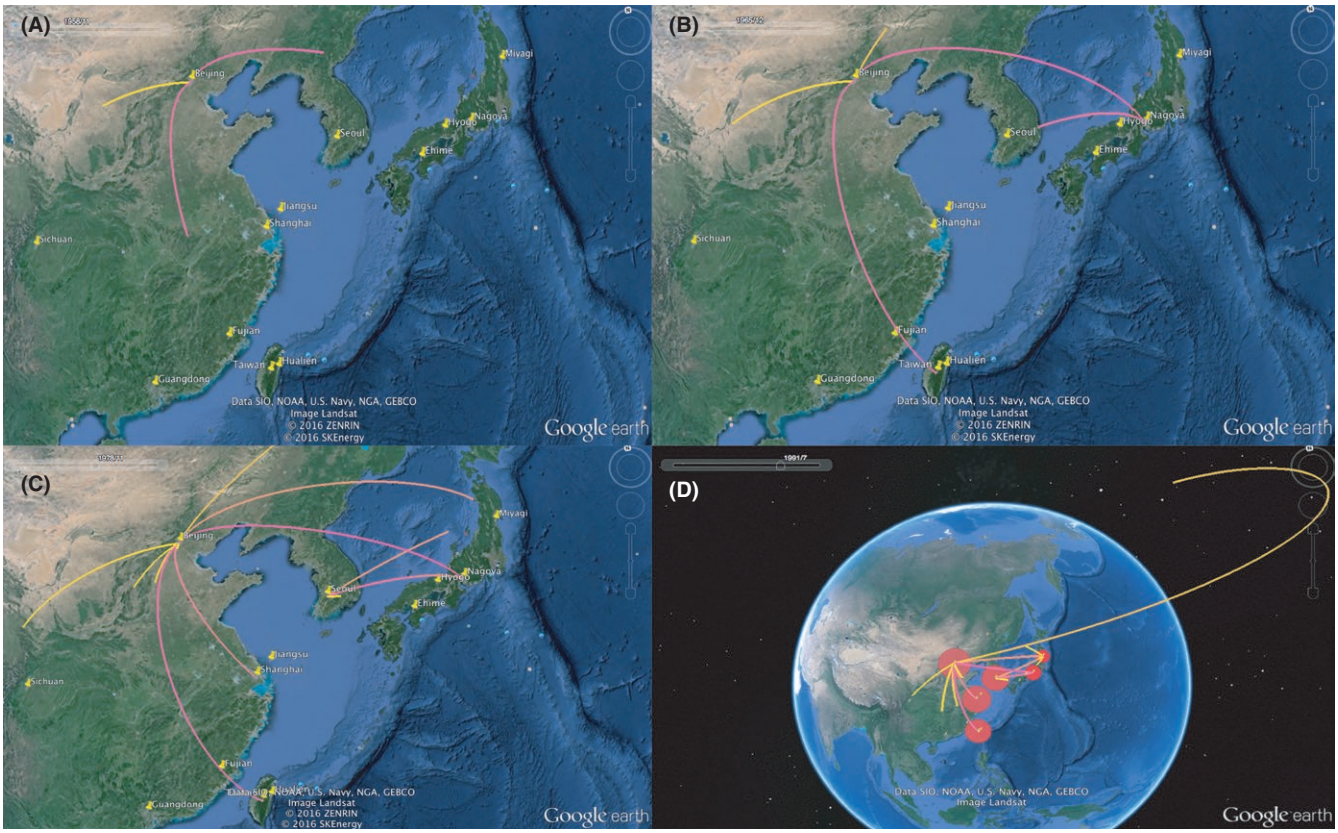

FIGURE 4 Temporal dynamics of HBV/C phylogeographic dispersion in East Asia. (A) Major routes of viral dispersion (pink line) set off from China to Japan and Taiwan respectively, while minor route (yellow line) started to expand inside China. (B) HBV lineage in Japan goes to Korea and flow back and forth (C). The direction and expansion range (circles in (D)) were inferred by the Bayesian MCC tree and visualized by Google earth (see Section 2)

Overall, in agreement with the ML and NJ gene flow analysis (see section 3.1), the data implicated China as the epicentre of several HBV/C radiations that were possibly at the origin of independent outbreaks in the adjacent Asian countries during the past three decades.

3.3 | HBV/C population dynamics and human mobility

Similar to the data set including all lineages, the relaxed molecular clock and nonparametric Skygrid demographic prior were determined to be the best-fitting models when sequences of each of the three major lineages described in the previous section were analysed separately (Table 1). Skygrid plots display changes in virus effective population size (*Ne*), which is a measure of viral diversity and is related to the number of new transmissions (incidence) over the sampling time period.[38] The population demographic history, or changes in *Ne* over the evolutionary history in question, of lineages 2 and 3 was remarkably similar (Figure 5). Lineage 3 was characterized by overall viral effective population growth from 1980 to 1997 (a small and transient decrease was observed between 1989 and 1995) (Figure 5A). A sharp population bottleneck was then observed in 2003, followed by growth and potential stabilization (growing HPD intervals toward most recent sampling time) during the last decade (Figure 5A). Such a strong bottleneck usually associated with one or more than two major events relate to natural environment, human or epidemics. Coincidently, a wide affected emerging pathogen outbreak arose in East Asia in 2003. This newly emerging coronavirus caused severe acute respiratory syndrome (SARS-CoV), leading to an outbreak in the area between Hong Kong and Guangdong in 2003. The fast transmission and high mortality rate caused worldwide attention. With the WHO and governments' propaganda continuously advising people to avoid the attendance of public events and to change travel plans to epidemic areas such as China and Taiwan,[39] restrictions on tourism were tightened so as to reduce human contact. In order to analyse the potential impact of these travel restrictions on HBV/C spread, we collected travel numbers at customs in each East Asia country during the past three decades (Figure 6) for comparison with phylogeographic data. The total entry number at customs arrival point of China and Taiwan (the two countries where most of lineage 3 strains circulated) showed a rapid growth from 1981 to 2002. In 2003, during the SARS-CoV epidemic, the number of arriving people had a loss of more than 6 million (97 908 252 people in 2002 vs 91 662 082 people in 2003) in China and 72 000 people in Taiwan (2 977 692 people in 2002 vs 2 248 117 people in 2003).

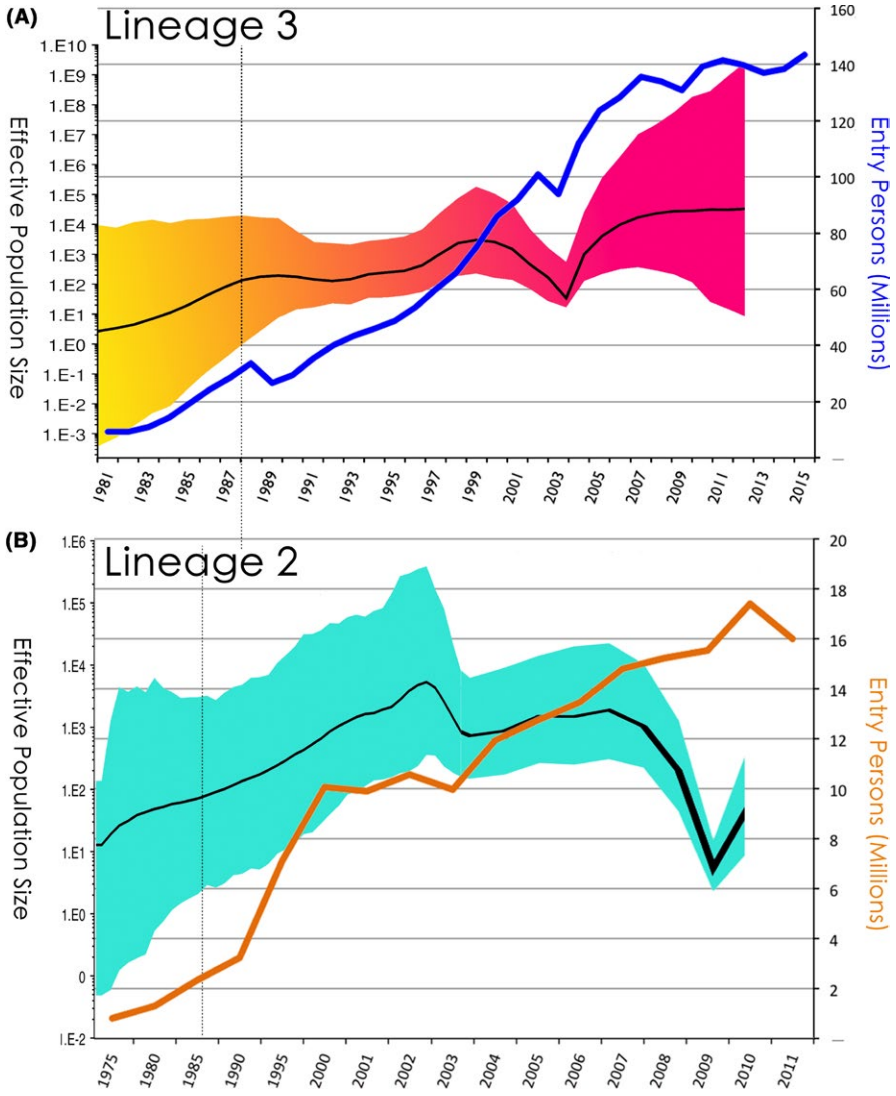

FIGURE 5 HBV/C Skygrid plots and human mobility data in East Asia. (A) The Bayesian Skygrid plot (yellow-pink) shows changes in viral effective population size (*Ne*) of HBV/C lineage 3 strains (left *y*-axis) over time (*x*-axis), with the black solid line and shaded areas representing mean N_e and 95% HPD estimates, respectively. The blue solid line represents human mobility data as the number of people entering China/Taiwan (right *y*-axis). (B) The Bayesian Skygrid plot (light blue colour) shows changes in viral *Ne* of HBV/C lineage 2 strains (left *y*-axis) over time (*x*-axis), with the black solid line and shaded areas representing mean *Ne* and 95% HPD estimates, respectively. The orange line represents human mobility data (right *y*-axis) as the number of people entering Japan. Human mobility data for Korea were not available

The population dynamic of lineage 3 is highly correlating with the number of people entering each country (Pearson $r = 0.829^{**}$, $P < 0.001$ for China; Pearson $r = 0.879^{**}$, $P < 0.001$, for Taiwan) and leaving each country (Pearson $r = 0.909^{**}$, $P < 0.001$ for China; Pearson $r = 0.709^{**}$, $P < 0.001$, for Taiwan). The HBV/C population bottleneck occurred during the time frame of the SARS-CoV epidemic[40] and was followed by another substantial growth matching a new increase in human mobility in the aftermath of the SARS-CoV epidemic (Figure 5).

Lineage 2 was characterized by fairly constant growth in viral *Ne* between 1975 and 2002, followed by two major bottlenecks (and eventual rebound): one in 2003 and a more recent one in 2009 (Figure 5B). The total entry number at customs arrival point of Japan and Korea (where lineage 2 strains mostly circulated) grew exponentially since 1975 and experienced a temporary decrease of 620 000 people in 2003 (10 585 117 people in 2002 vs 9 964 762 people in 2003), followed by another rapid growth and eventual plateau in 2004-2008 (Figure 5B) in the aftermath of the SARS-CoV emergency. Comparison with human mobility data shows a significant correlation between the viral *Ne* and number of people travelling out of Japan and Korea (Pearson $r = 0.578^{*}$, $P = 0.038$), as well as with people entering Japan and Korea (Pearson $r = 0.555^{*}$, $P = 0.049$), while the viral *Ne* population bottleneck in 2003 matches the reduced travel out of Japan that occurred during the SARS-CoV epidemic scare.

4 | DISCUSSION

The prevalence and dynamics of HBV/C infection are a major public health concern in East Asia. Our study resulted in two significant findings. First, we showed that during the past three decades HBV/C appeared to have been disseminated radically from China to neighbouring Eastern Asian countries—through multiple independent introductions and further spread such as, for example, from Japan to Korea—all phylogeographic transitions corresponded with human mobility. This is not only evident in the directional spread from China/Taiwan to Japan and from Japan to Korea, but also by the introduction of HBV/C strains in USA from China around the 1960s shown by our phylogeography analysis. The correlation of

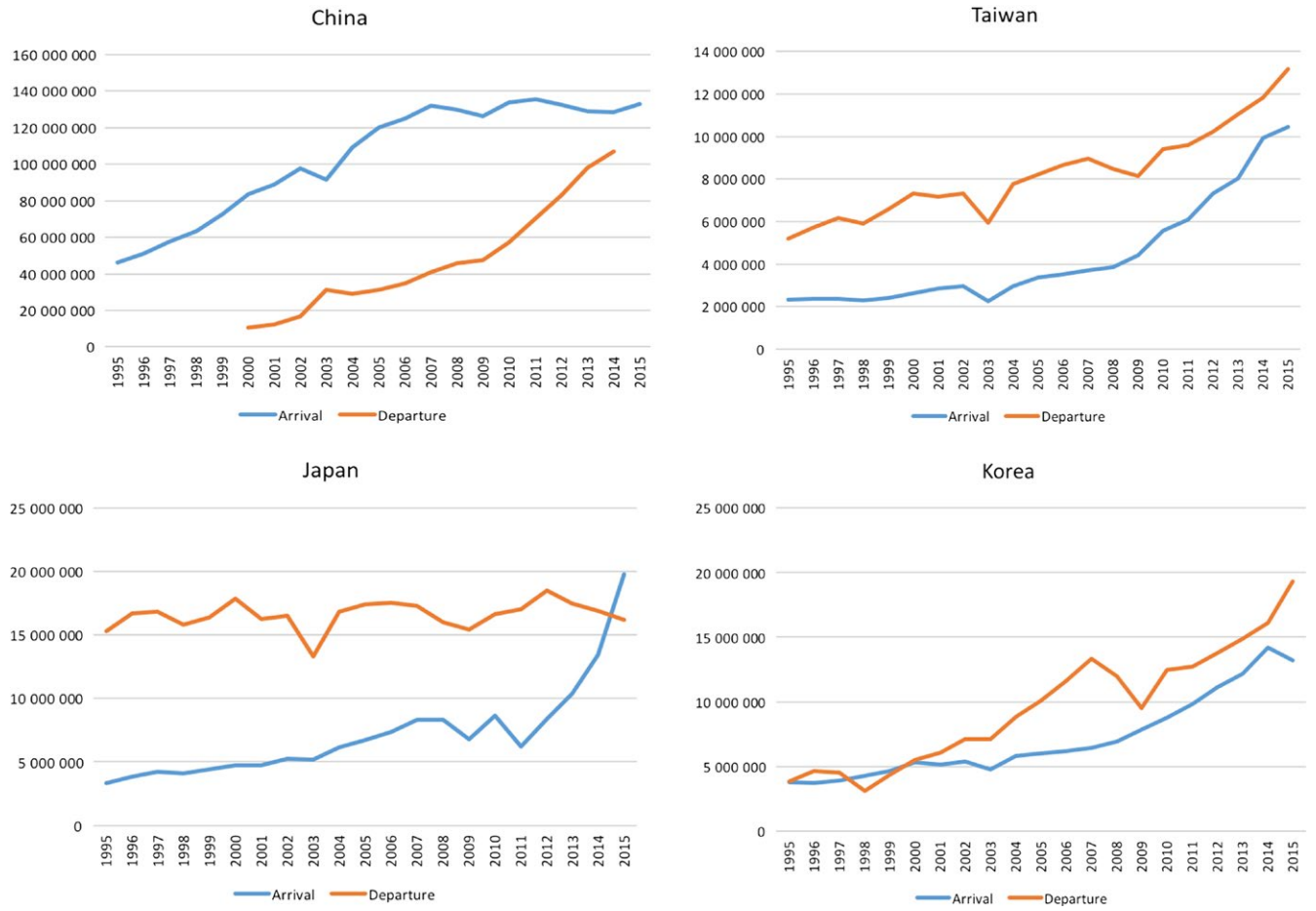

FIGURE 6 Arriving and departure numbers in East Asia. The orange pattern line represents departure number in each year (*x*-axis) and the blue line represents the arriving number at the custom check point in China, Taiwan, Japan and Korea, respectively. Persons were plotted by millions (*y*-axis)

dissemination of virus out of China with neighbouring countries over the past three decades with human mobility indicates a prominent role for human travel in spreading disease in East Asia. In 1966, the "Proletarian Cultural Revolution" in China sparked public demonstrations, factional battles and government initiatives that resulted in waves of immigration to the USA,[41,42] exactly the same time frame of HBV/C introduction, according to our reconstruction.

The strong HBV population bottleneck is confusing at first glance because there seems no appropriate explanation for any event leading to the major decline in East Asia around that time. However, SARS-CoV raised an outbreak in East Asia. It led to 774 deaths and 8098 infections and reached 30 countries in only a few months.[43] The fast transmission and high mortality rate (9.6%)[44] caused worldwide attention. During the 2003 SARS outbreak, hospitals with SARS cases and healthcare workers had been segregated.[45] Thousands of people had been isolated in hospitals including medical staff. People who had the possibility of being in contact with SARS patients were requested to be isolated in their homes.[46,47] Under this social atmosphere, community fairs and gatherings were reduced, not only in the general population but also in PWID (patients with injecting drugs) and MSM (men having sex with men). Human contact and international population flow (eg, travel) are important factors in the genetic diversity supply to a blood-transmitted virus such as HBV. When the effective population lacks income for a while, this would then be sharply reflected in phylodynamic analysis. Human-contact-caused environmental pressure for virus *Ne* is too hard to prove, but the pressure has left a trace on travelling numbers in every East Asia country.

While the overall *Ne* of the virus has been increasing over the past three decades, at least two of the major co-circulating lineages of genotype C (one mainly spreading in Japan and Korea, the other one in China and Taiwan) experienced a significant viral population bottleneck around 2003 at the time of the SARS-CoV outbreak in China. The analysis of human mobility data showed that, indeed, during the SARS-CoV emergency, population movements in East Asia were strongly correlated with viral *Ne* fluctuation. The relatively limited decrease in the number of foreigners entering Japan and Korea may be due to the absence of Japan and Korea in the list of WHO travel advice recipients, also potentially explaining the moderate statistical correlation with the *Ne* of lineage 2. The significant decrease in people leaving the country during 2003, together with the phylogeography reconstruction showing directional

HBV/C flow from Japan to Korea, provides a possible explanation for the temporary decrease in viral *Ne*. A similar conclusion could be drawn for the 2003 viral population bottleneck in the lineage mainly spreading in China with multiple introductions to Taiwan, which correlated with a reduced number of people entering Taiwan from China (155 872 people in 2002 vs 133 422 people in 2003).

Since 1995, when successful vaccination programs started to curtail the number of "mother-to-infant" transmissions, overall HBV prevalence has been decreasing, especially in Japan.[48] However, current vaccination programs do not target recent immigrants from neighbouring countries, and our data show a steady increase in viral diversity, consistent with continuous emergence and dissemination of distinct lineages, likely fostered by human migration.

HBV/C and SARS-CoV infect human hosts through very different routes (airborne vs mainly sexual contact), but human mobility reflected the influence on the host of co-expansion of HBV from the spread of SARS-CoV. In this particular case, the restriction on human mobility caused by the rapid dissemination of SARS-CoV may have resulted in a temporary decrease in the HBV/C effective population size. In other words, our data indicate that the co-circulation of two pathogens targeting the same human population resulted in a complex dynamic wherein one outbreak effectively down-modulated the other. A stochastic model of pathogen competition, co-circulating in a spatially structured environment, has recently been proposed.[49] By simulating two pathogens spreading through a travelling network with a stochastic mechanistic approach, the model investigated several complex scenarios where "ecological factors," such as host mobility, could be either noninfluential for the competition dynamics or play a critical role in selecting the dominant pathogen. The development of more realistic spatial ecology models of multiple, co-circulating infectious diseases, as well as the monitoring and analysis of potentially competing epidemics, should be, therefore, a major focus for next-generation epidemiology studies.

This is the first study depicting the evolutionary history of HBV/C in East Asia. By using state-of-the-art phylogeography techniques, we show that distinct HBV epidemic lineages emerged from China during the 1930s-1960s and spread exponentially across different Eastern Asian countries in the following decades. Demographic inference of viral effective population size over time indicated similar dynamics for both lineages, characterized by exponential growth since the early 1980s, followed by a significant bottleneck in 2003 and another increase after 2004. We first speculated that the host co-expansion virus population was influenced by strict restriction of population movement from/to China during the SARS-CoV outbreak in 2003. Given the continuous challenge of emerging and re-emerging pathogens in the 21st century globalized world, our results highlight the need for the development of more realistic spatial ecology models of multiple co-circulating infectious diseases.

ACKNOWLEDGEMENTS

This study was supported in part by the Ministry of Science and Technology, Taiwan (grant NSC 102-2917-I-019-001), and the Mackay Memorial Hospital, Taiwan (grant MMH-106-125 and 2017-C01-00152).

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ORCID

Hsin-Fu Liu http://orcid.org/0000-0003-0082-2269

REFERENCES

1. World Health Organization. WHO guidance on development of influenza vaccine reference viruses by reverse genetics; 2005.
2. Okamoto H, Tsuda F, Sakugawa H, et al. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol*. 1988;69(10):2575-2583.
3. Utsumi T, Yano Y, Truong BX, et al. Molecular epidemiological study of hepatitis B virus infection in two different ethnic populations from the Solomon Islands. *J Med Virol*. 2007;79(3):229-235.
4. Kramvis A, Kew M, Francois G. Hepatitis B virus genotypes. *Vaccine*. 2005;23(19):2409-2423.
5. Chu CJ, Keeffe EB, Han SH, et al. Hepatitis B virus genotypes in the United States: results of a nationwide study. *Gastroenterology*. 2003;125(2):444-451.
6. Tatematsu K, Tanaka Y, Kurbanov F, et al. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol*. 2009;83(20):10538-10547.
7. Osiowy C, Kaita K, Solar K, Mendoza K. Molecular characterization of hepatitis B virus and a 9-year clinical profile in a patient infected with genotype I. *J Med Virol*. 2010;82(6):942-948.
8. Cho JH, Yoon KH, Lee KE, et al. Distribution of hepatitis B virus genotypes in Korea. *Korean J Hepatol*. 2009;15(2):140-147.
9. Chan HL, Wong GL, Tse CH, et al. Hepatitis B virus genotype C is associated with more severe liver fibrosis than genotype B. *Clin Gastroenterol Hepatol*. 2009;7(12):1361-1366.
10. Orito E, Ichida T, Sakugawa H, et al. Geographic distribution of hepatitis B virus (HBV) genotype in patients with chronic HBV infection in Japan. *Hepatology*. 2001;34(3):590-594.
11. Zeng G, Wang Z, Wen S, et al. Geographic distribution, virologic and clinical characteristics of hepatitis B virus genotypes in China. *J Viral Hepat*. 2005;12(6):609-617.
12. Kao JH, Chen PJ, Lai MY, Chen DS. Genotypes and clinical phenotypes of hepatitis B virus in patients with chronic hepatitis B virus infection. *J Clin Microbiol*. 2002;40(4):1207-1209.
13. Huy TT, Ushijima H, Quang VX, et al. Genotype C of hepatitis B virus can be classified into at least two subgroups. *J Gen Virol*. 2004;85(Pt 2):283-292.
14. Sugauchi F, Mizokami M, Orito E, et al. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. *J Gen Virol*. 2001;82(Pt 4):883-892.
15. Chan HL, Wong ML, Hui AY, Hung LC, Chan FK, Sung JJ. Hepatitis B virus genotype C takes a more aggressive disease course than hepatitis B virus genotype B in hepatitis B e antigen-positive patients. *J Clin Microbiol*. 2003;41(3):1277-1279.
16. Health promotion administration MoHaW; 2014. http://www.hpa.gov.tw/BHPNet/Portal/file/FormCenterFile/20141024094922044/online_english/index.html - 80.

17. Chan HL, Tsang SW, Liew CT, et al. Viral genotype and hepatitis B virus DNA levels are correlated with histological liver damage in HBeAg-negative chronic hepatitis B virus infection. *Am J Gastroenterol*. 2002;97(2):406-412.
18. Kumar A, Kumar SI, Pandey R, Naik S, Aggarwal R. Hepatitis B virus genotype A is more often associated with severe liver disease in northern India than is genotype D. *Indian J Gastroenterol*. 2005;24(1):19-22.
19. Lindh M, Hannoun C, Dhillon AP, Norkrans G, Horal P. Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. *J Infect Dis*. 1999;179(4):775-782.
20. Screening for chronic hepatitis B among Asian/Pacific Islander populations–New York City, 2005. MMWR - Morbidity & Mortality Weekly Report; 2006. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5518a2.htm.
21. Zhou Y, Holmes EC. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J Mol Evol*. 2007;65(2):197-205.
22. Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. Molecular evolution of hepatitis B virus over 25 years. *J Virol*. 2006;80(21):10307-10314.
23. Guirgis BS, Abbas RO, Azzazy HM. Hepatitis B virus genotyping: current methods and clinical implications. *Int J Infect Dis*. 2010;14(11):e941-e953.
24. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725-2729.
25. Liang X, Bi S, Yang W, et al. Epidemiological serosurvey of hepatitis B in China–declining HBV prevalence due to hepatitis B vaccination. *Vaccine*. 2009;27(47):6550-6557.
26. Juon HS, Choi KS, Park EC, Kwak MS, Lee S. Hepatitis B vaccinations among Koreans: Results from 2005 Korea National Cancer Screening Survey. *BMC Infect Dis*. 2009;9:185.
27. Miyamoto T. Epidemiology of Hepatitis B Virus in Japan, especially in Nagasaki. *Acta Med Nagasaki*. 1986;30:290-293.
28. Chen CJ, Wang LY, Yu MW. Epidemiology of hepatitis B virus infection in the Asia-Pacific region. *J Gastroenterol Hepatol*. 2000;15(Suppl):E3-E6.
29. Kim WR. Epidemiology of hepatitis B in the United States. *Hepatology*. 2009;49(5 Suppl):S28-S34.
30. Swofford DL. http://paup.phylosolutions.com. Accessed September 21, 2018.
31. Hudson RR, Slatkin M, Maddison WP, Estimation of levels of gene flow from DNA sequence data. *Genetics*. 1992;132(2):583-589.
32. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol*. 2013;30(3):713-724.
33. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 2012;29(9):2157-2167.
34. Baele G, Lemey P, Vansteelandt S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics*. 2013;14:85.
35. Kass RERAE. Bayes Factor. *J Am Stat Assoc*. 1995;90:773-795.
36. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29(8):1969-1973.
37. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*. 2011;27(20):2910-2912.
38. Frost SD, Volz EM. Viral phylodynamics and the search for an 'effective number of infections'. *Philos Trans R Soc Lond B Biol Sci*. 2010;365(1548):1879-1890.
39. WHO. 2003 SARS epidemic area http://www.who.int/csr/sars/travel/en/.
40. Salemi M, Fitch WM, Ciccozzi M, Ruiz-Alvarez MJ, Rezza G, Lewis MJ. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *J Virol*. 2004;78(3):1602-1603.
41. Geng Z. *Cultural Revolution memoirs written and read in English: Image formation, reception and counternarrative:* Thesis(Ph.D.)-University of Minnesota; 2008.
42. Cheng N. *Life and Death in Shanghai*. New York: Grove Press; 1986.
43. WHO. World Health Organization changes Hong Kong, Guangdong travel recommendations; 2003. http://www.who.int/csr/sars/travel/en/
44. WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003; 2003. http://www.who.int/csr/sars/country/table 2003_09_23/en/. Accessed September 26, 2003.
45. McDonald LC, Simor AE, Su IJ, et al. SARS in healthcare facilities, Toronto and Taiwan. *Emerg Infect Dis*. 2004;10(5):777-781.
46. Twu SJ, Chen TJ, Chen CJ, et al. Control measures for severe acute respiratory syndrome (SARS) in Taiwan. *Emerg Infect Dis*. 2003;9(6):718-720.
47. Hsieh Y-H, Chen CWS, Hsu S-B. SARS Outbreak, Taiwan, 2003. *(CDC) EID J*. 2004;10:201-206. https://wwwnc.cdc.gov/eid/article/10/2/03-0515_article. Accessed September 21, 2018.
48. Lin SY, Toyoda H, Kumada T, Liu HF. Molecular evolution and phylodynamics of acute hepatitis B virus in Japan. *PLoS ONE*. 2016;11(6):e0157103.
49. Poletto C, Meloni S, Van Metre A, Colizza V, Moreno Y, Vespignani A. Characterising two-pathogen competition in spatially structured environments. *Sci Rep*. 2015;5:7895.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lin SYC, Magalis BR, Salemi M, Liu H-F. 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. https://doi.org/10.1111/jvh.13006