The genetic diversity and evolutionary history of hepatitis C virus in Vietnam

Chunhua Li[1], **Manqiong Yuan**[1,†], **Ling Lu**[1,*], **Teng Lu**[1,2,¶], **Wenjie Xia**[3], **Van H. Pham**[4,5], **An X.D. Vo**[5], **Mindie H. Nguyen**[6], and **Kenji Abe**[4,5,7,*]

[1] Department of Pathology and Laboratory Medicine, Center for Viral Oncology, University of Kansas Medical Center, Kansas City, Kansas, United States of America

[2] University of Southern California, Los Angeles, California, United States of America

[3] Guangzhou Blood Center, Guangzhou, Guangdong, China

[4] Center for Molecular Biomedicine, School of Medicine, University of Medicine and Pharmacy in Ho Chi Minh City, Ho Chi Minh City

[5] Molecular Diagnostics Section, The NK-BIOTEK Laboratory, Ho Chi Minh City, Vietnam

[6] Division of Gastroenterology and Hepatology, Stanford University Medical Center, Palo Alto, California

[7] Department of Pathology, National Institute of Infectious Diseases, Tokyo 162–8640, Japan

Abstract

Vietnam has a unique history in association with foreign countries, which may have resulted in multiple introductions of the alien HCV strains to mix with those indigenous ones. In this study, we characterized the HCV sequences in Core-E1 and NS5B regions from 236 Vietnamese individuals. We identified multiple HCV lineages; 6a, 6e, 6h, 6k, 6l, 6o, 6p, and two novel variants may represent the indigenous strains; 1a was probably introduced from the US; 1b and 2a possibly originated in East Asia; while 2i, 2j, and 2m were likely brought by French explorers. We inferred the evolutionary history for four major subtypes: 1a, 1b, 6a, and 6e. The obtained Bayesian Skyline Plots (BSPs) consistently showed the rapid HCV population growth from 1955-1963 until 1984 or after, corresponding to the era of the Vietnam War. We also estimated

[*] Corresponding author: Mailing address: Department of Pathology, National Institute of Infectious Diseases, Toyama 1 23 1, Shinjuku ku, Tokyo 162 8640, Japan. Tel: +81-3-4582-2701, Fax: +81-3-5285-1189, kenjiabe@nih.go.jp Correspondence: Ling Lu, Center for Viral Oncology, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, MSN 3046, 3901 Rainbow Boulevard, Kansas City, KS 66160. llu@kumc.edu. Tel: 913-945-6451 Fax: 913-945-6836..
[†]Current address: State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, School of Public Health, Xiamen University, Xiamen, Fujian, 361102, PR China.
[¶]Summer Student for Bioinformatics.

HCV growth rates and reconstructed phylogeographic trees for comparing subtypes 1a, 1b, and HCV-2.

Keywords

HCV; genotype; sequence; BEAST; evolution; Vietnam

INTRODUCTION

Hepatitis C virus (HCV) is one of the major causative agents of chronic liver disease. Globally, an estimated 3% of the world population are chronically infected with the virus (Alter, 2007). Chronic HCV infection significantly increases the risk of liver cirrhosis and hepatocellular carcinoma, which are associated with substantial morbidity and mortality that are expected to increase over the next decades (Jahan et al., 2012). According to a report from the World Health Organization (WHO, 2012), countries with high HCV prevalence include those in Africa and Asia (Perz *et al.*, 2006; Shepard *et al.*, 2005; Sievert *et al.*, 2011) such as Egypt (14.7%) (Guerra *et al.*, 2012), Pakistan (4.95%) (Waheed *et al.*, 2009), Thailand (3.2%-5.6%) (Apichartpiyakul *et al.*, 1999; Songsivilai *et al.*, 1997), and Vietnam (2%-9%) (Nakata *et al.*, 1994; Tran *et al.*, 2003).

Analysis of sequences has classified HCV into seven genotypes (Smith et al., 2014). Except for genotypes 5 and 7, each of them is further divided into multiple subtypes. Different genotypes and subtypes of HCV have shown unique patterns of geographic distribution. In general, subtypes 1a, 1b, 2a, 2b, and 3a are globally distributed. In contrast, genotype 4 is primarily limited to indigenous populations in North Africa and the Middle East, 5a to South Africa, and genotype 6 to Southeast Asia and IDUs and expatriates from that region (Pybus *et al.*, 2009; Pybus *et al.*, 2007). HCV genotype is an important indicator for patient management because its genetic variation patterns are associated with responses to the therapy of pegylated interferon plus ribavirin (Ahmad *et al.*, 2011; Fried *et al.*, 2002).

Recently, a large scale of HCV survey has been completed in Vietnam, which screened 8,654 individuals with different risk factors in five geographic regions: Ha Noi, Hai Phong, Da Nang, Khanh Hoa, and Can Tho. This study determined the HCV genotypes among 282 individuals and found that subtypes 1a (33%) and 1b (27%) were the most common followed by 6a (18.8%), while other genotype 6 subtypes such as 6e (6.0%), 6h (4.6%), and 6l (6.4%) were also frequent (Dunford *et al*., 2012). In contrast, subtypes 2a (0.4%), 3a (0.7%), and 3b (1.1%) were rare. In that study, however, neither were included the individuals from Ho Chi Minh City, nor were the evolutionary patterns analyzed. Therefore, the aim of the current article was to estimate the HCV epidemic history and evolutionary dynamics based on the sequences determined from the HCV infected individuals sampled in Ho Chi Minh City, the largest population center in Vietnam and the one with the most possible exposure to the US personnel during the Vietnam War. It is hoped that the yielded results may add new insights to the current understanding of the HCV infection and transmission in the country and the given evolutionary estimates may be more informative than the data obtained using the traditional approaches.

MATERIALS AND METHODS

Subjects and specimens

Serum samples were obtained from 236 individuals infected with HCV. These individuals included blood donors and patients with liver disease. They were recruited by the Cho Ray Hospital, the Hospital for Tropical Diseases, and the University of Medicine and Pharmacy, all in Ho Chi Minh City, Vietnam. Informed consent for participation in this study was obtained from each individual. The ethical review committees of the above three institutes in Vietnam and the University of Kansas Medical Center in USA and the National Institute of Infectious Diseases in Japan approved this study.

Sequence amplification and genotyping

The amplified Core-E1 and NS5B sequences of HCV correspond to the nucleotide numbering 739-1310 and 8267-8630 in the H77 genome. These sequences were determined as previously described (Fu *et al.*, 2011; Lu *et al.*, 2005) followed by co-analyses with 13 reference sequences representing 13 subtypes that were detected in this study. These references were selected based on the recommendations in a recent paper by Smith et al. (2014). Prior to phylogenetic analysis, the best-fitting substitution model was tested using the jModeltest program (Posada, 2008) according to the Akaike Information Criterion, which indicated that GTR+I+Γ was the best model for both Core-E1 and NS5B datasets. Under this model, maximum likelihood (ML) trees were heuristically searched using the SPR (Subtree Pruning and Regrafting) and NNI (Nearest Neighbor Interchange) algorithms in PhyML (Guindon & Gascuel, 2003). With the tree files generated and applied to the Figtree program (version 1.4) (Pybus & Rambaut, 2009), tree topology was displayed and converted into a circular form. To characterize the novel genotype 6 variants, 67 references were co-analyzed (see Supplementary data for their selections). Based on these resulting datasets, recent virus recombination events were excluded using the RDP3 program under the settings as previously described (Lu *et al.*, 2008).

Evolutionary analysis

HCV evolutionary history was estimated for subtypes 1a, 1b, 6a, and 6e, using the Markov Chain Monte Carlo (MCMC) algorithm in the BEAST package (version 1.7.1). Briefly, the best combination of the GTR+I+Γ substitution model, the Bayesian skyline coalescent model, and the uncorrelated exponential clock model was used on the basis that it always outperforms other combinations (Drummond *et al.*, 2012; Fu *et al.*, 2012; Pybus *et al.*, 2009), after a formal model test that showed the ln Bayes factors ranged between 1.98-25.82. However, we included different evolutionary rates as priors. Through directly estimating the sequences determined in this study using the Path-o-gen software (http://tree.bio.ed.ac.uk/software/pathogen), we obtained four such rates, 9.1025×10^{-4} substitution/site/year for 1a, -0.6×10^{-3} for 1b, 1.163×10^{-3} for 6a, and 1.1646×10^{-3} for 6e. Because the 1b rate was negative, it was discarded. Instead, we used a rate of 1.02×10^{-3} that we have recently estimated, based on an alternative 1b dataset (Yuan *et al.,* 2013). To allow for adequate statistical uncertainties, we set a credible interval at 1×10^{-5} substitution/site/year in the BEAST analyses.

To explore the phylogeographic nature of 1a, 1b, 2a, and 2m sequences resulting from this study, additional BEAST analyses were performed after the addition of a number of reference sequences followed by testing using Befi-BaTS (www.lonelyjoeparker.com/?tag=befi-bats). The references included 176 1a, 166 1b, 52 2a, and 46 HCV-2 sequences (see Supplementary data for their selections).

To compare the differences in the patterns of HCV population growth among subtypes, HCV growth rates were estimated within the periods of rapid growth and continuous growth. We first exported a BSP log file from the Tracer to obtain the median number of HCV effective population sizes and then used this information in a simple piecewise linear regression analysis, from which the generated slope measures the speed of HCV population growth. To show the obvious switches of growth curve, we performed a natural logarithmic transformation to the number of HCV effective population sizes and re scaled the BSP for inspection. To divide the periods of rapid growth and continuous growth mathematically, we started two regression analyses on both sides of a certain time point and then slid this procedure through all the time points in the exported BSP log file within a certain range, which allowed the regression analyses to minimally span 10 time points. Such a sliding produced a curve of sum of r^2 that was used to identify the optimum time point at which the sum of r^2 maximized. It indicates that the two regression analyses divided by this time point are the best-fitting among all. For simplicity, we wrote an R script for each BSP to run all of these procedures automatically (R Core Team, 2013). To obtain more precise HCV growth rates with their 95% confidence intervals, we ran the BEAST program under a "parametric" constant logistic model as recently described (Lu *et al.*, 2013).

After the above parameters were set using the BEAUti program, XML files were generated and applied to the BEAST program (Fu *et al.*, 2012; Yuan *et al.*, 2013). The latter ran MCMC procedures each for 300 million states and output a tree every 10,000 states. To assess the MCMC sampling convergence, the estimated effective sampling sizes (ESS) were evaluated. In this study, when all of the ESS numbers were   200, sufficient sampling was considered to have been achieved. To interpret the MCMC chains and output posterior trees, the Tracer program (version 1.5) was used. To generate phylogeographic trees in a decreasing node order, the resulting posterior tree files were deciphered using the Figtree program (version 1.4).

RESULTS

Phylogenetic analysis

Core-E1 and NS5B sequences of HCV were determined in 236 subjects: 146 (61.86%) men, 86 (36.44%) women, and four with unknown gender. All of them were Vietnamese and aged from 22 to 88, with a mean age of 46.13 ± 11.5. Figure 1 and 2 presented two ML trees in circular form, reconstructed using the obtained Core-E1 and NS5B sequences and co-analyzed with 13 reference sequences representing 13 assigned subtypes that were related to this study. Both trees revealed considerable genetic diversity of HCV representing six prevalent subtypes: 1a in 29 (12.3%), 1b in 48 (20.3%), 2a in 20 (8.5%), 2m in 12 (5.1%), 6a in 53 (22.5%), and 6e in 48 (20.3%). In addition, seven uncommon subtypes were also detected: 6h in 2, 6k in 3, 6l in 6, 6o in 4, 6p in 7, 2i and 2j each in one. Furthermore, two

isolates, TV469 and VN214, were classified under HCV-6, but their subtype could not be determined (Table 1 and Figure 1). As a whole, HCV-6 isolates accounted for more than a half (53%). Consistently shown in both trees, sequences of identical isolates were located in similar positions and therefore, reliable sequencing results were confirmed and no recent viral recombination or mixed HCV infection was suggested. In both trees, when a subtype cluster was formed, a significant bootstrap support of >81% was shown.

A possible new subtype

TV469 and VN214 may represent a new subtype. For verification, 67 genotype 6 reference sequences were retrieved from GenBank and co-analyzed, which generated additional two trees: one for the Core-E1 (Figure S1) and the other for NS5B (Figure S2). Both trees showed TV469 and VN214 forming a close subset to distinct from all other HCV-6 lineages. In Figure S1, we also included 48 subtype 6e sequences determined in this study and they composed an extended but diverse cluster. Interestingly, TV469 and VN214 appeared to be more relative to 6e than to any other subtypes. Recently, partial NS5B sequences have been characterized for many novel HCV-6 variants, which were all sampled in Vietnam (Koletzki *et al.*, 2010) and also co-analyzed in Figure S2 (blue branches). To a certain extent, five of these sequences were more relative to TV469 and VN214. Among them, GU049374, TV469, and VN214 appeared to form a closer lineage to loosely group with GU049368, GU049385, GU049389, and GU049400 in forming a larger subset. Via GU049381 as an intermediate variant, this larger subset was adjacent to the 6e clade and as a whole they may represent an ancient subtype that was historically endemic in Vietnam.

Based on Figures S1-S2, pairwise nucleotide similarities were calculated. In the Core-E1 and NS5B regions, TV469 and VN214 shared similarities of 94.1-94.8% to each other, suggesting different strains of a single subtype. Comparing with the 67 HCV-6 references, TV469 and VN214 showed similarities of 67.1-79.3% in the Core-E1 and 68.1-79.4% in the NS5B region, while comparing with GU049374 in NS5B, the similarities were 81.7-83.4%. These similarities indicate different subtypes.

Evolutionary analysis

Based on the Core-E1 sequences of 1a, 1b, 6a, and 6e that each had a percentage of more than 10% among the detected 236 isolates, four evolutionary analyses were performed. The first analysis used the 29 1a sequences with a prior rate of 9.1×10^{-4}. A BSP was given that showed the median growth curve steadily increasing from 1963 until 1990.8 (Figure 3, Table 3). Although this growth curve became slightly slower after 1990.8, it remained to go up until the present. Since considerable boundaries were included, a wide range of the 95% highest posterior density confidence intervals (HPD) was given, particularly from 1963 to 1980. The second analysis used the 48 sequences of 1b with a prior rate 1.02×10^{-3}. The resulting BSP showed the median growth curve going up from 1961 until 1985.1, yet after 1985.1 the growth slowed to the present. The third analysis used the 53 sequences of 6a with a prior rate 1.116×10^{-3}. The resulting BSP showed an almost identical pattern to that of 1b. However, the median growth curve indicated the rapid HCV growth starting before 1958 up to 1984.2. The fourth analysis used the 48 sequences of 6e with a prior rate 1.1646×10^{-3}. The resulting BSP showed a standard "S" shape of the median growth curve. In the latter

BSP, a relatively constant population size was maintained before 1925 till 1950 that was followed by an exponential population growth that continued to 1986.2 and then slowed to the present.

Phylogeographic tree analysis

Phylogeographic trees for 1a and 1b showed slightly different patterns. A noted feature for 1a is that the Vietnamese sequences diverged between the mid-1950s and early 1960s, while many US and a few European 1a clusters emerged later (Figure S3). For 1b, such time points, at which the Vietnamese sequences emerged, were in a much wider range and appeared to be continual, from the earliest in 1941 to the latest after 1971 (Figure S4). In addition, the Vietnamese 1b clusters had more complicated internal branches than those of 1a. Collectively, these would suggest more active divergences and perhaps a longer duration of circulation of the Vietnamese 1b than 1a isolates.

After classifying the 1a sequences into three groups (US, Europe, and Vietnam) and 1b into four (US, Europe, China, and Vietnam), both trees were tested for two tree shape statistics, AI (Association Index) and PS (Parsimony Score), which showed values strongly rejecting the null hypothesis of panmix (Parker *et al.*, 2008; Slatkin & Maddison, 1989). Thus, the existence of the geographic clusters is strongly supported. In fact, these geographic clusters can be steadily seen in both trees, particularly for 1a, in which three Vietnamese clusters were separable from those of the US and Europe. Vietnamese clusters were also seen in the tree for 1b, although many sequences from the US and Europe appeared to mix. Based on these Vietnamese 1a and 1b geographic clusters, the most recent common ancestors (tMRCAs) were estimated by additional BEAST analyses, which showed the times of origin in 1950 (95% HPD: 1938, 1960), 1947 (95% HPD: 1934, 1961), and 1958 (95% HPD: 1947, 1968) for three 1a clusters, and 1948 (95% HPD: 1936, 1959), 1937 (95 HPD: 1918, 1953), and 1940 (95% HPD: 1928, 1953) for three 1b clusters, with the emergences of 1b earlier than 1a.

Similarly, phylogeographic trees were reconstructed for subtypes 2a and 2m as well (Figures S5-S6), but the results were described and discussed in the supplementary data.

Growth rate estimation

The four BSPs generated using Tracer may not show the clear patterns of growth curve for indicating different growth periods (Fig. 3). However, after a natural logarithmic transformation of the median population sizes exported from the BSP log files, such patterns appeared to be obvious (Fig.4). Based on the transformed growth curves, two growth periods, the rapid growth and the continuous growth, were mathematically divided at the time points where the largest sum of r^2 of regression was seen (Table 3). This resulted that the rapid growth was delimited from 1963 to 1990.8 for 1a, from 1961 to 1985.1 for 1b, from 1958 to 1984.2 for 6a, and from 1949.8 to 1986.2 for 6e, while the continuous growth was synchronized for the four subtypes after 1990 to the present (Fig. 3). Using simple linear regression to analyze the median population sizes within these two growth periods, growth rates were estimated that are equal to the regression slopes (Fig. 4). Within the period of rapid growth, the slopes (slope 1) were 0.087, 0.193, 0.182, and 0.136 for 1a, 1b,

6a, and 6e, respectively. Within the period of continuous growth, the slopes (slope 2) were 0.034, 0.041, 0.039, and 0.029 (Table 3). Using slope 1 as a prior to further estimate the same datasets by running the BEAST program under the constant logistic model (Pybus *et al.*, 2003), more precise HCV growth rates (mean ± 95% confidence intervals) were estimated within the period of rapid growth: 0.2295 (95%CI: 0.1351, 0.343) for 1a, 0.3581 (95%CI: 0.212, 0.5203) for 1b, 0.3001 (95%CI: 0.184, 0.4233) for 6a, and 0.2494 (95%CI: 0.1386, 0.3702) for 6e (Table 3).

DISCUSSION

In this study, we characterized the HCV sequences from 236 individuals in Ho Chi Minh City. Phylogenetic analysis revealed a high degree of genetic diversity, confirming the polyphyletic nature of HCV in Vietnam. Thirteen subtypes 1a, 1b, 2a, 2i, 2j, 2m, 6a, 6e, 6h, 6k, 6l, 6o, 6p, and two novel genotype 6 variants were characterized with sequences congruently determined from two separated genomic regions. Among them, genotypes 1 and 6 were predominant, accounting for 32.63% (77/236) and 52.97% (125/236), respectively. Although genotype 2 was at 18.64% (44/236), some isolates were exclusively detectable in Vietnam. Different from other studies in Southeast Asian countries (Hubschen *et al.*, 2011; Jutavijittum *et al.*, 2009; Shinji *et al.*, 2004; Sistayanarain *et al.*, 2011) including those from Vietnam (Dunford L, 2012; Pham *et al.*, 2009; Tanimoto *et al.*, 2010), the proportions of 1a, 1b, 6a, and 6e were compatible while no genotype 3 was detectable, consistent with a recent report (Pham *et al.*, 2011). These isolates constituted a unique HCV genotype distribution pattern, characteristic of a long term endemic circulation mixed with subsequent multiple introductions of the alien strains.

The first finding is a high percentage of HCV-6 isolates (53%, 125/236) and among them 6a is predominant accounting for 42.4% (53/125). Although 6a was first discovered in Hong Kong (Adams *et al.*, 1997; Bukh *et al.*, 1992), more divergent 6a isolates were later sequenced from blood donors in Ho Chi Minh City (Tokita *et al.*, 1994). Currently, 6a infections are common in Thailand, Myanmar, and China, and also detectable among Asian immigrants in North America. Recently, we studied the 6a migration patterns in China and included many 6a sequences from Vietnam as references that were determined from the cohort reported here (Fu *et al.*, 2012).

Among the detected HCV-6 isolates, 6e was the second predominant (38.4%, 48/125). Two features were noted for 6e: (1) genetically highly diverse and complex; (2) often detected in Vietnam and occasionally in the neighboring China, Thailand, and Cambodia, and in Asian immigrants in north America as well (Akkarathamrongsin *et al.*, 2010, 2011; Murphy *et al.*, 2007; Noppornpanth *et al.*, 2006; Pham *et al.*, 2009; Tokita *et al.*, 1994). According to these features, we believe that 6e has been long indigenous to Vietnam and subsequently spread to the neighboring countries. For example, via the known drug trafficking routes, it has spread and become locally epidemic in Guangxi province in China bordering Vietnam (Garten *et al.*, 2005).

Subtypes 6h, 6l, 6k, 6o, and 6p were also detected. In addition, two variants TV496 and VN214 were identified that may qualify for a new subtype. With the inclusion of subtypes

6b, 6d, 6j, 6q, 6t, and many unclassified variants archived in GenBank, a wide spectrum of HCV-6 members can be represented, indicating a long endemic circulation and perhaps, the earliest origin in Vietnam.

New HCV-6 variants continue to be identified and the majority were detected in Vietnam (Lu *et al.*, 2008; Murphy *et al.*, 2007). A proposal for a unified HCV nomenclature has detailed the guidelines for classifying a new subtype. Minimally, it requires three closely related isolates independently identified with sequences differing from known subtypes by 15% of nucleotides in Core/E1 and NS5B regions, while the three should group consistently using different methods (Simmonds *et al.*, 2005; Smith et al., 2014). According to this premise, both VN214 and TV469 may represent a new subtype, since they had nucleotide similarities of 94.1-94.8% in Core-E1 and NS5B regions while differing from other HCV-6 lineages by nucleotides of 16.6%-32.9%. Furthermore, they showed no recent viral recombination events. However, since they lack the minimal three closely related isolates detectable, the provisional designation of a new subtype is premature at the moment.

The second finding is that subtype 1a accounted for 12.3% (29/236). Although this is lower (*P*<0.05) than 23.7%-33.0% recently reported for 1a detected among various patients from other regions in Vietnam (Dunford *et al.*, 2012; Pham *et al.*, 2009; Tanimoto *et al.*, 2010), such a rate is still high. We are not certain if this was predicted by different compositions of the study subjects who, at different ages or with varied risk factors, may show preference for different HCV subtypes. However, due to the lack of patient information, we were unable to assess this. Subtype 1a represents the major strain in the US (51.6%) (Nainan et al., 2006). In contrast, it is rare in China (2.1%) (Fu et al., 2011), Myanmar (4.1%) (Lwin et al., 2007), and Cambodia (2.3%, our unpublished data), and has been never reported in Laos. Regardless, it is common in Thailand (14.3%) (Jutavijittum et al., 2009) and more common in the Philippines (34%-36%) (Katayama et al., 1996; Maramag et al., 2006). Between the latter two countries, there appears to be an association of 1a prevalence with the interactions from the US. A relatively high frequency of 1a in Vietnam may have reflected such interactions during the Vietnam War.

The third finding is a relatively low proportion of subtype 1b (20.34%, 48/236). This subtype accounts for approximately 66% of the HCV infections in China (Lu et al., 2005), 64.4% in Japan, and 66% in South Korea (Ohno et al., 1997). It is also one of the major subtypes in Europe (Esteban et al., 2008).

A number of unique HCV-2 isolates were detected. These included 12 isolates of subtype 2m and one each of 2i and 2j. Recently, we summarized the HCV-2 sequences from GenBank (Li *et al.*, 2012). We noted that 2m had origins exclusively in Vietnam (Fig.S6) (Newman *et al.*, 2012), while 2i and 2j were not only detectable in Vietnam but also in France and other countries, such as 2i in Germany, Canada, and Morocco (Li *et al.*, 2012) and 2j in Spain (Holland *et al.*, 1996) and Venezuela (Sulbaran *et al.*, 2010). According to these data, we may hypothesize that all HCV-2 strains had their ancestral origins in Africa (Li *et al.*, 2012). However, they were brought outside Africa most likely by European explorers. Since Vietnam was a French colony over a century, these HCV-2 strains were likely introduced via exposure to foreign nationals. This is consistent with the recent

findings that many HCV-2 strains were detected in Indonesia, Canada, France, Martinique, and the island of Hispaniola. Subtype 2m has been only found in Vietnam but never in other countries, which may be due to a limited sampling in Africa, particularly in those countries that are also former French colonies (Sulbaran *et al.*, 2010).

Four BSPs were generated. They consistently showed the rapid HCV population growth from about 1950-1963 until 1984 or after. Largely, this time frame corresponded to the Vietnam War and the Third Indochina War (http://www.ushistory.org/us/55.asp). The former war started in 1954, escalated in the early 1960s with the US involvement in South Vietnam, and ended in April 1975, marked by the fall of Saigon, which is now called Ho Chi Minh City, where all of the study subjects for this study were recruited. However, the latter war occurred only in the North and was in a small-scale and short-lived, which started and ended in 1979.

The median growth curve of the 1a-BSP did not show a clear separation of different phases. Instead, it exhibited a continuous growth in HCV-infected population size, though with a slight slowing around 1990. The beginning of 1a-BSP in 1963 may correspond to the time point when 1a strains, likely originating in the US, were introduced to Vietnam and rapidly added to the number of new cases during wartime. The 1a growth was maintained even after the war, likely because this alien strain has been firmly seeded and then become epidemic among the local population. Such a presumption may explain why the first half of the median growth curve of the 1a-BSP presented a pattern similar to that was previously estimated for the 1a growth in the US (Nakano *et al.*, 2004). However, there could be additional reasons for the continuous growth represented by the second half of the median growth curve that remained to rise even after 1990, although in a slower slope (S2=0.034). In a previous report, a marked slowing in the increase in the 1a-infected population size in the US was also seen around 1990, which has been attributed to the discovery of HCV in 1989 and a subsequent implementation of HCV screening assays in 1990 (Nakano *et al.*, 2004). In war-time Vietnam, a large number of people were wounded and many may have been infected with the alien 1a strains due to exposure to blood products and medical services provided by foreign personnel. In post-war Vietnam, HCV-contaminated blood transfusions remained a serious problem until a policy of voluntary unpaid blood donation was enforced by the government in recent years, which has finally resulted in a steady rise in the proportion of blood donors who have passed the mandatory screening for HIV, HBV, and HCV (http://www.who.int/mediacentre/factsheets/fs279/en/index.html). Regardless, a reservoir of 1a-infected people, founded during the war time, continued to spread the infection in the general population and contributed to the continuous increase in the incidences of HCV transmissions, as presented by the current 1a-BSP.

In comparison, the start of 1b-BSP appeared to be slightly earlier than 1a and the start of 6a earlier than 1b (Fig.3). This is consistent with the fact that 1b represents the major subtype in Asia and 6a is a local epidemic lineage in Vietnam (Dunford et al, 2012; Pham *et al.*, 2009; Tanimoto *et al.*, 2010). Phylogeographic tree analysis further showed that the Vietnamese 1b clusters had older tMRCAs and more complicated internal branches than the Vietnamese 1a clusters. This result corroborates a previous finding that the global expansion of 1b preceded that of 1a by at least 16 years (Magiorkinis et al., 2009). Based on the

analyses in this study, it may be thought that 1b and also 6a represent the pre-existing HCV lineages sharing very similar BSP patterns, and both grew at faster rates than that seen with subtype 1a (Slope 1: 1b=0.193, 6a=0.182, 1a=0.087), especially during war time. However, following the end of the war, 1984 and after, the growth became much slower, yet they remained on the rise to the present time, at similar rates among subtypes (Slope 2: 1b=0.041, 6a=0.039, 1a=0.034). Compared with 1a, 1b, and 6a, the 6e growth was preceded with an extended period of constant population size from 1930 till 1950, while the pattern of its rapid growth during the war time was highly similar to that of 1b and 6a. Hence, 6e is characteristic of a long-term endemic circulation with a rapid growth during the war. The indigenous circulation of 6e is also supported by the fact that the 6e sequences are genetically diverse (Figures S1-S2) and almost exclusively determined in Vietnam (http://hcv.lanl.gov/).

The BSPs for 1a and 1b both showed wide ranges of errors from their beginning to about 1985 (Fig.3). One of the potential meanings of such profiles is that, under the used prior rates and credible intervals, the hypothesis of a constant size population cannot be ruled out. In other words, because the molecular clock information is not conclusive enough, these profiles may suggest the lack of HCV fast population growth. With no knowledge about how much these errors reflected the statistical uncertainties, nor obtaining the information about the dates when the HCV infections had been acquired by those individuals from whom we sampled the 1a and 1b sequences, and the unavailability of the historical data to show the overtime changes of HCV prevalence in the general population in Vietnam, we are now difficult to draw any conclusion that is free from controversies. However, the fast growth of 1a and 1b can be suggested by findings from other studies in two aspects. In the first aspect, based on the 1a and/or 1b sequences sampled globally or in countries, patterns of fast HCV growth have been indicated. These patterns largely corresponded to what we estimated in this study, although to a certain extent their starting and ending dates varied (Nakano et al., 2004; Lu et al., 2013; Magiorkinis et al., 2009). Based on a short NS5B region sequenced, similar BSP patterns have been also estimated for the 1a and 1b strains sampled in Vietnam. The results showed that the 1a growth started around 1960s, while the 1b growth was slower but started earlier in 1920 (Nakano et al., 2004). In another study at the global level, robust estimates have been provided; they indicated the massive expansion of 1a and 1b strains between 1940 and 1980. While the earliest divergences of 1a and 1b occurred in developed countries, the subsequent spread to developing countries appeared more often in the recent (Magiorkinis et al., 2009). From the latter study, the estimated time frame implied the global transmissions of 1a and 1b being likely initiated and sustained by the increase of parenteral iatrogenic procedures during and after World War II, for which blood and lyophilized pooled plasma could have served as important vehicles (Power et al., 1995). Via these vehicles, large numbers of 1a and 1b transmissions could have also occurred during the Korean War (Kendrick, 1964) and the Vietnam War (Neel, 1991). The fast growth of 1a and 1b also coincided with the history of illicit IDU. While in the developed countries the heroin use by means of IDU peaked in 1960s, such an activity is relatively recent in many Asian countries including Vietnam (Stimson, 1993). Since 1a is more related to IDU (Rosen et al., 1999) while 1b is preferentially transmitted by transfusions and unsafe medical injections (Dubois et al., 1997; Pawlotsky et al., 1995), it

may be speculated that a considerable portion of the 1a introductions into Vietnam were mediated via IDU, particularly by the US soldiers during the war. Of the studies in the second aspect, the earliest one reported an anti-HCV prevalence of 9% among the inhabitants in Ho Chi Minh City who were not with any liver disease (Nakata et al., 1994). In 2003, a much lower rate of 2% was revealed among the healthy people in the same city (Tran et al., 2003). Nevertheless, both rates were higher than 1% that was later reported in another study based on the randomly sampled general population in a rural area in proximity to the city (Nguyen et al., 2007). These rates may be not comparable, because they were from different reports. However, they may have reflected a temporal decreasing trend of HCV prevalence in the recent decades in Ho Chi Minh City where we had collected samples for this study. Such a trend is consistent with our BSP estimates that showed lower HCV growth since 1990 and after, and hence implies a fast HCV growth period before 1990.

Previous studies have reported a series of HCV growth rates during the period of rapid growth: 0.25 for genotype 4 in Egypt (Pybus *et al.*, 2003), 0.1 for 1a and 3a in the US and UK (Pybus *et al.*, 2005; Tanaka *et al.*, 2006), and 0.17 for 6a in Hong Kong (Tanaka *et al.*, 2006). In this study, we estimated the median growth rates for the four major HCV subtypes in Vietnam during their periods of rapid growth, using a parametric model in BEAST. This gave the 1b rate (0.3499) higher than 6a (0.2929), then 6e (0.2416), and then 1a (0.2232). The 6a and 6e rates are both higher than that of 6a in Hong Kong, and the 1a rate higher than that of 1a in the US. Although they may somehow reflect the increased incidences of HCV transmission during the 20-year-long Vietnam War, these rates were not as high as those we have recently reported in China (Lu *et al.*, 2013). Although the Vietnam War was believed to be the significant historical event for founding the currently seen high HCV prevalence in Vietnam, the induced incidences of HCV transmission appeared to be moderate. This may be due to the lack of a marked rise in birth rate and the many casualties caused by the war that resulted in a smaller population size at risk for HCV infection. However, in the postwar era a great level of HCV transmissions was maintained in Vietnam, as represented by the growth rates estimated during this period for the four HCV subtypes: 0.034 for 1a, 0.041 for 1b, 0.039 for 6a, and 0.032 for 6e. It is implied that some common reasons for HCV transmission and dissemination have not been removed, which requires additional measures taken, not only for preventing HCV transmission but also for improving the general condition of public health.

It should be noted that the effective population size was represented as base 10 logarithm in the four BSPs, and it was also represented so in the "piecewise logistic" population model. However, we used regression to produce the slopes (growth rates) after a natural logarithmic transformation for the purpose to show more obvious switches of the growth curves for the mathematical division of the growth periods. Therefore, the estimated two growth rates, one by regression analysis and the other by using a parametric model, are not directly comparable. Further extrapolation of these estimates should consider their differences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

Adams NJ, Chamberlain RW, Taylor LA, Davidson F, Lin CK, Elliott RM, Simmonds P. Complete coding sequence of hepatitis C virus genotype 6a. Biochem Biophys Res Comm. 1997; 234:393–396. [PubMed: 9177282]

Ahmad W, Ijaz B, Javed FT, Kausar H, Sarwar MT, Gull S, Asad S, Shahid I, Hassan S. HCV genotype specific correlation with serum markers: higher predictability for genotype 4a. Virol J. 2011; 8:293. [PubMed: 21663629]

Akkarathamrongsin S, Praianantathavorn K, Hacharoen N, Theamboonlers A, Tangkijvanich P, Tanaka Y, Mizokami M, Poovorawan Y. Geographic distribution of hepatitis C virus genotype 6 subtypes in Thailand. J Med Virol. 2010; 82:257–262. [PubMed: 20029811]

Akkarathamrongsin S, Praianantathavorn K, Hacharoen N, Theamboonlers A, Tangkijvanich P, Poovorawan Y. Seroprevalence and genotype of hepatitis C virus among immigrant workers from Cambodia and Myanmar in Thailand. Intervirology. 2011; 54:10–16. [PubMed: 20689311]

Alter MJ. Epidemiology of hepatitis C virus infection. World J Gastroenterol. 2007; 13:2436–2441. [PubMed: 17552026]

Apichartpiyakul C, Apichartpiyakul N, Urwijitaroon Y, Gray J, Natpratan C, Katayama Y, Fujii M, Hotta H. Seroprevalence and subtype distribution of hepatitis C virus among blood donors and intravenous drug users in northern/northeastern Thailand. Jpn J Infect Dis. 1999; 52:121–123. [PubMed: 10507992]

Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5' noncoding region of hepatitis C virus. Proc Nat Acad Sci USA. 1992; 89:4942–4946. [PubMed: 1317578]

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29:1969–1973. [PubMed: 22367748]

Dubois F, Desenclos JC, Mariotte N, Goudeau A. Hepatitis C in a French population-based survey, 1994: seroprevalence, frequency of viremia, genotype distribution, and risk factors. The Collaborative Study Group. Hepatology. 1997; 25:1490–1496. [PubMed: 9185773]

Dunford L, Carr MJ, Dean J, Waters A, Nguyen LT, Ta Thi TH, Thi LA, Do HD, Thi TT, Nguyen HT, Diem Do TT, Luu QP, Connell J, Coughlan S, Nguyen HT, Hall W,W, Nguyen Thi LA. Hepatitis C Virus (HCV) in Vietnam: High Prevalence of Infection in Dialysis and Multi-transfused Patients Involving Diverse and Novel Virus Variants. PloS one. 2012; 7:e41266. [PubMed: 22916104]

Esteban JI, Sauleda S, Quer J. The changing epidemiology of hepatitis C virus infection in Europe. J Hepatol. 2008; 48:148–162. [PubMed: 18022726]

Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL Jr. Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. Peginterferon alfa 2a plus ribavirin for chronic hepatitis C virus infection. New Engl J Med. 2002; 347:975–982. [PubMed: 12324553]

Fu Y, Qin W, Cao H, Xu R, Tan Y, Lu T, Wang H, Tong W, Rong X, Li G, Yuan M, Li C, Abe K, Lu L, Chen G. HCV 6a prevalence in Guangdong province had the origin from Vietnam and recent dissemination to other regions of China: phylogeographic analyses. PloS one. 2012; 7:e28006. [PubMed: 22253686]

Fu Y, Wang Y, Xia W, Pybus OG, Qin W, Lu L, Nelson K. New trends of HCV infection in China revealed by genetic analysis of viral sequences determined from first time volunteer blood donors. J Viral Hepat. 2011; 18:42–52. [PubMed: 20196805]

Garten RJ, Zhang J, Lai S, Liu W, Chen J, Yu XF. Coinfection with HIV and hepatitis C virus among injection drug users in southern China. Clin Infect Dis. 2005; 41(Suppl 1):S18–24. [PubMed: 16265609]

Guerra J, Garenne M, Mohamed MK, Fontanet A. HCV burden of infection in Egypt: results from a nationwide survey. J Viral Hepat. 2012; 19:560–567. [PubMed: 22762140]

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]

Holland PV, Barrera JM, Ercilla MG, Yoshida CF, Wang Y, de Olim GA, Betlach B, Kuramoto K, Okamoto H. Genotyping hepatitis C virus isolates from Spain, Brazil, China, and Macau by a simplified PCR method. J Clin Microbiol. 1996; 34:2372–2378. [PubMed: 8880482]

Hubschen JM, Jutavijittum P, Thammavong T, Samountry B, Yousukh A, Toriyama K, Sausy A, Muller CP. High genetic diversity including potential new subtypes of hepatitis C virus genotype 6 in Lao People's Democratic Republic. Clin Microbiol Infect. 2011; 17:E30–34. [PubMed: 21958219]

Jahan S, Ashfaq UA, Qasim M, Khaliq S, Saleem MJ, Afzal N. Hepatitis C virus to hepatocellular carcinoma. Infect Agent Cancer. 2012; 7:2. [PubMed: 22289144]

Jutavijittum P, Jiviriyawat Y, Yousukh A, Pantip C, Maneekarn N, Toriyama K. Genotypic distribution of hepatitis C virus in voluntary blood donors of northern Thailand. Southeast Asian J Trop Med Pub Health. 2009; 40:471–479. [PubMed: 19842432]

Katayama Y, Barzaga NG, Alipio A, Soetjipto, Doi H, Ishido S, Hotta H. Genotype analysis of hepatitis C virus among blood donors and inmates in Metro Manila, The Philippines. Microbiol Immunol. 1996; 40:525–529. [PubMed: 8865159]

Kendrick, D. Blood program in World War II. Supplemented by experiences in the Korean War. Boyd Coates J, MC., editor. Office of the Surgeon General Department of the Army; Washington (D.C.): 1964.

Koletzki D, Dumont S, Vermeiren H, Fevery B, De Smet P, Stuyver LJ. Development and evaluation of an automated hepatitis C virus NS5B sequence-based subtyping assay. Clin Chem Lab Med. 2010; 48:1095–1102. [PubMed: 20578969]

Li C, Cao H, Lu L, Murphy D. Full-length sequences of 11 hepatitis C virus genotype 2 isolates representing five subtypes and six unclassified lineages with unique geographical distributions and genetic variation patterns. J Gen Virol. 2012; 93:1173–1184. [PubMed: 22357752]

Lu L, Murphy D, Li C, Liu S, Xia X, Pham PH, Jin Y, Hagedorn CH, Abe K. Complete genomes of three subtype 6t isolates and analysis of many novel hepatitis C virus variants within genotype 6. J Gen Virol. 2008; 89:444–452. [PubMed: 18198375]

Lu L, Nakano T, He Y, Fu Y, Hagedorn CH, Robertson BH. Hepatitis C virus genotype distribution in China: predominance of closely related subtype 1b isolates and existence of new genotype 6 variants. J Med Virol. 2005; 75:538–549. [PubMed: 15714489]

Lu L, Tong W, Gu L, Li C, Lu T, Tee KK, Chen G. The current HCV prevalence in China may have mainly resulted from an officially encouraged plasma campaign in the 1990s: a coalescence inference with genetic sequences. J Virol. 2013; 87:12041–12050. [PubMed: 23986603]

Lwin AA, Shinji T, Khin M, Win N, Obika M, Okada S, Koide N. Hepatitis C virus genotype distribution in Myanmar: Predominance of genotype 6 and existence of new genotype 6 subtype. Hepatol Res. 2007; 37:337–345. [PubMed: 17441806]

Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, Pybus OG, Allain JP, Hatzakis A. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Med. 2009; 6:e1000198. [PubMed: 20041120]

Maramag F, Rivera M, Predicala R, Baclig M, Matias R, Cervantes J. Hepatitis C genotypes among Fillipinos. Phil J Gastroenterol. 2006; 2:30–32.

Markov PV, van de Laar TJ, Thomas XV, Aronson SJ, Weegink CJ, van den Berk GE, Prins M, Pybus OG, Schinkel J. Colonial history and contemporary transmission shape the genetic diversity of hepatitis C virus genotype 2 in Amsterdam. J Virol. 2012; 86:7677–7687. [PubMed: 22573865]

Murphy DG, Willems B, Deschenes M, Hilzenrat N, Mousseau R, Sabbah S. Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5' untranslated region sequences. J Clin Microbiol. 2007; 45:1102–1112. [PubMed: 17287328]

Nakano T, Lu L, Liu P, Pybus OG. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. J Infect Dis. 2004; 190:1098–1108. [PubMed: 15319860]

Nakata S, Song P, Duc DD, Nguyen XQ, Murata K, Tsuda F, Okamoto H. Hepatitis C and B virus infections in populations at low or high risk in Ho Chi Minh and Hanoi, Vietnam. J Gastroenterol Hepatol. 1994; 9:416–419. [PubMed: 7524723]

Nainan OV, Alter MJ, Kruszon-Moran D, Gao FX, Xia G, McQuillan G, Margolis HS. Hepatitis C virus genotypes and viral concentrations in participants of a general population survey in the United States. Gastroenterology. 2006; 131:478–484. [PubMed: 16890602]

Neel, S. The Military Blood Program. Medical support of the U.S. army in Vietnam 1965-1970. Spurgeon, Neel, editor. Office of Medical History, Department of the Army; Washington (D.C.): 1991.

Newman RM, Kuntzen T, Weiner B, Berical A, Charlebois P, Kuiken C, Murphy DG, Simmonds P, Bennett P, Lennon NJ, Birren BW, Zody MC, Allen TM, Henn MR. Whole genome pyrosequencing of rare hepatitis C virus genotypes enhances subtype classification and identification of naturally occurring drug resistance variants. J Infect Dis. 2012; 208:17–31. [PubMed: 23136221]

Noppornpanth S, Lien TX, Poovorawan Y, Smits SL, Osterhaus AD, Haagmans BL. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. J Virol. 2006; 80:7569–7577. [PubMed: 16840336]

Nguyen VT, McLaws ML, Dore GJ. Prevalence and risk factors for hepatitis C infection in rural north Vietnam. Hepatology international. 2007; 1:387–393. [PubMed: 19669334]

Ohno O, Mizokami M, Wu RR, Saleh MG, Ohba K, Orito E, Mukaide M, Williams R, Lau JY. New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. J Clin Microbiol. 1997; 35:201–207. [PubMed: 8968908]

Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. Infect Genet Evol. 2008; 8:239–246. [PubMed: 17921073]

Pawlotsky JM, Tsakiris L, Roudot Thoraval F, Pellet C, Stuyver L, Duval J, Dhumeaux D. Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. J Infect Dis. 1995; 171:1607–1610. [PubMed: 7769300]

Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J Hepatol. 2006; 45:529–538. [PubMed: 16879891]

Pham DA, Leuangwutiwong P, Jittmittraphap A, Luplertlop N, Bach HK, Akkarathamrongsin S, Theamboonlers A, Poovorawan Y. High prevalence of Hepatitis C virus genotype 6 in Vietnam. Asian Pacific J Allergy Immunol. 2009; 27:153–160.

Pham VH, Nguyen HD, Ho PT, Banh DV, Pham HL, Pham PH, Lu L, Abe K. Very high prevalence of hepatitis C virus genotype 6 variants in southern Vietnam: large-scale survey based on sequence determination. Jpn J Infect Dis. 2011; 64:537–539. [PubMed: 22116339]

Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008; 25:1253–1256. [PubMed: 18397919]

Power JP, Lawlor E, Davidson F, Holmes EC, Yap P,L, Simmonds P. Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti D immunoglobulin. Lancet. 1995; 345:1211–1213. [PubMed: 7739308]

Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, Syhavong B, Phetsouvanah R, Sheridan I, Humphreys IS, Lu L, Newton PN, Klenerman P. Genetic history of hepatitis C virus in East Asia. J Virol. 2009; 83:1071–1082. [PubMed: 18971279]

Pybus OG, Cochrane A, Holmes EC, Simmonds P. The hepatitis C virus epidemic among injecting drug users. Infect Genet Evol. 2005; 5:131–139. [PubMed: 15639745]

Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol. 2003; 20:381–387. [PubMed: 12644558]

Pybus OG, Markov PV, Wu A, Tatem AJ. Investigating the endemic transmission of the hepatitis C virus. Inter J Parasitol. 2007; 37:839–849.

Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009; 10:540–550. [PubMed: 19564871]

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013. URL http://www.R-project.org/

Rosen HR, Chou S, Sasaki AW, Gretch DR. Molecular epidemiology of hepatitis C infection in US veteran liver transplant recipients: evidence for decreasing relative prevalence of genotype 1B. Am J Gastroenterol. 1999; 94:3015–3019. [PubMed: 10520862]

Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis. 2005; 5:558–567. [PubMed: 16122679]

Shinji T, Kyaw YY, Gokan K, Tanaka Y, Ochi K, Kusano N, Mizushima T, Fujioka S, Shiraha H, Lwin AA, Shiratori Y, Mizokami M, Khin M, Miyahara M, Okada S, Koide N. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. Acta medica Okayama. 2004; 58:135–142. [PubMed: 15471435]

Sievert W, Altraif I, Razavi HA, Abdo A, Ahmed EA, Alomair A, Amarapurkar D, Chen CH, Dou X, El Khayat H, Elshazly M, Esmat G, Guan R, Han KH, Koike K, Largen A, McCaughan G, Mogawer S, Monis A, Nawaz A, Piratvisuth T, Sanai FM, Sharara AI, Sibbel S, Sood A, Suh DJ, Wallace C, Young K, Negro F. A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt. Liver international. 2011; 31(Suppl 2):61–80. [PubMed: 21651703]

Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin IT, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005; 42:962–973. [PubMed: 16149085]

Sistayanarain A, Kunthalert D, Vipsoongnern Y. A shift in the hepatitis C virus genotype dominance in blood donor samples from Thailand. Mol Biol Rep. 2011; 38:4287–4290. [PubMed: 21113670]

Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 1989; 123:603–613. [PubMed: 2599370]

Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P. Expanded classification of hepatitis C Virus into 7 genotypes and 67 Subtypes : updated criteria and assignment web resource. Hepatology. 2014; 59:318–327. [PubMed: 24115039]

Songsivilai S, Jinathongthai S, Wongsena W, Tiangpitayakorn C, Dharakul T. High prevalence of hepatitis C infection among blood donors in northeastern Thailand. Am J Trop Med Hyg. 1997; 57:66–69. [PubMed: 9242321]

Stimson GV. The global diffusion of injecting drug use: implications for human immunodeficiency virus infection. Bull Narc. 1993; 45:3–17. [PubMed: 8305905]

Sulbaran MZ, Di Lello FA, Sulbaran Y, Cosson C, Loureiro CL, Rangel HR, Cantaloube JF, Campos RH, Moratorio G, Cristina J, Pujol FH. Genetic history of hepatitis C virus in Venezuela: high diversity and long time of evolution of HCV genotype 2. PloS one. 2010; 5:e14315. [PubMed: 21179440]

Tanaka Y, Kurbanov F, Mano S, Orito E, Vargas V, Esteban JI, Yuen MF, Lai CL, Kramvis A, Kew MC, Smuts HE, Netesov SV, Alter HJ, Mizokami M. Molecular tracing of the global hepatitis C virus epidemic predicts regional patterns of hepatocellular carcinoma mortality. Gastroenterology. 2006; 130:703–714. [PubMed: 16530512]

Tanimoto T, Nguyen HC, Ishizaki A, Chung PT, Hoang TT, Nguyen VT, Kageyama S, Oka S, Pham VT, Ichimura H. Multiple routes of hepatitis C virus transmission among injection drug users in Hai Phong, Northern Vietnam. J Med Virol. 2010; 82:1355–1363. [PubMed: 20572071]

Tokita H, Okamoto H, Tsuda F, Song P, Nakata S, Chosa T, Iizuka H, Mishiro S, Miyakawa Y, Mayumi M. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. Proc Nat Acad Sci USA. 1994; 91:11022–11026. [PubMed: 7972001]

Tran HT, Ushijima H, Quang VX, Phuong N, Li TC, Hayashi S, Xuan Lien T, Sata T, Abe K. Prevalence of hepatitis virus types B through E and genotypic distribution of HBV and HCV in Ho Chi Minh City, Vietnam. Hepatol Res. 2003; 26:275–280. [PubMed: 12963426]

Tucker, SC. French Indo China. Tucker, Spencer C., editor. Vietnam.The Uiniversity Press of Kentucky; 1999. p. 22-25.

Waheed Y, Shafi T, Safi SZ, Qadri I. Hepatitis C virus in Pakistan: a systematic review of prevalence, genotypes and risk factors. World J Gastroenterol. 2009; 15:5647–5653. [PubMed: 19960560]

WHO. Hepatitis C fact sheet N°164.. 2012. Edited by http://www.who.int/mediacentre/factsheets/fs164/en/index.html

Yuan M, Lu T, Li C, Lu L. The evolutionary rates of HCV estimated with subtype 1a and 1b sequences over the ORF length and in different genomic regions. PloS one. 2013; 8:e64698. [PubMed: 23762247]

- We characterized both Core-E1 and NS5B sequences of HCV from 236 Vietnamese.

- We detected 13 subtypes plus two novel variants as a new subtype candidate.

- They represent multiple indigenous circulations as well as several alien strains.

- We estimated the epidemic history for four major subtypes: 1a, 1b, 6a, and 6e.

- They all showed rapid HCV growth corresponding to the era of the Vietnam War.
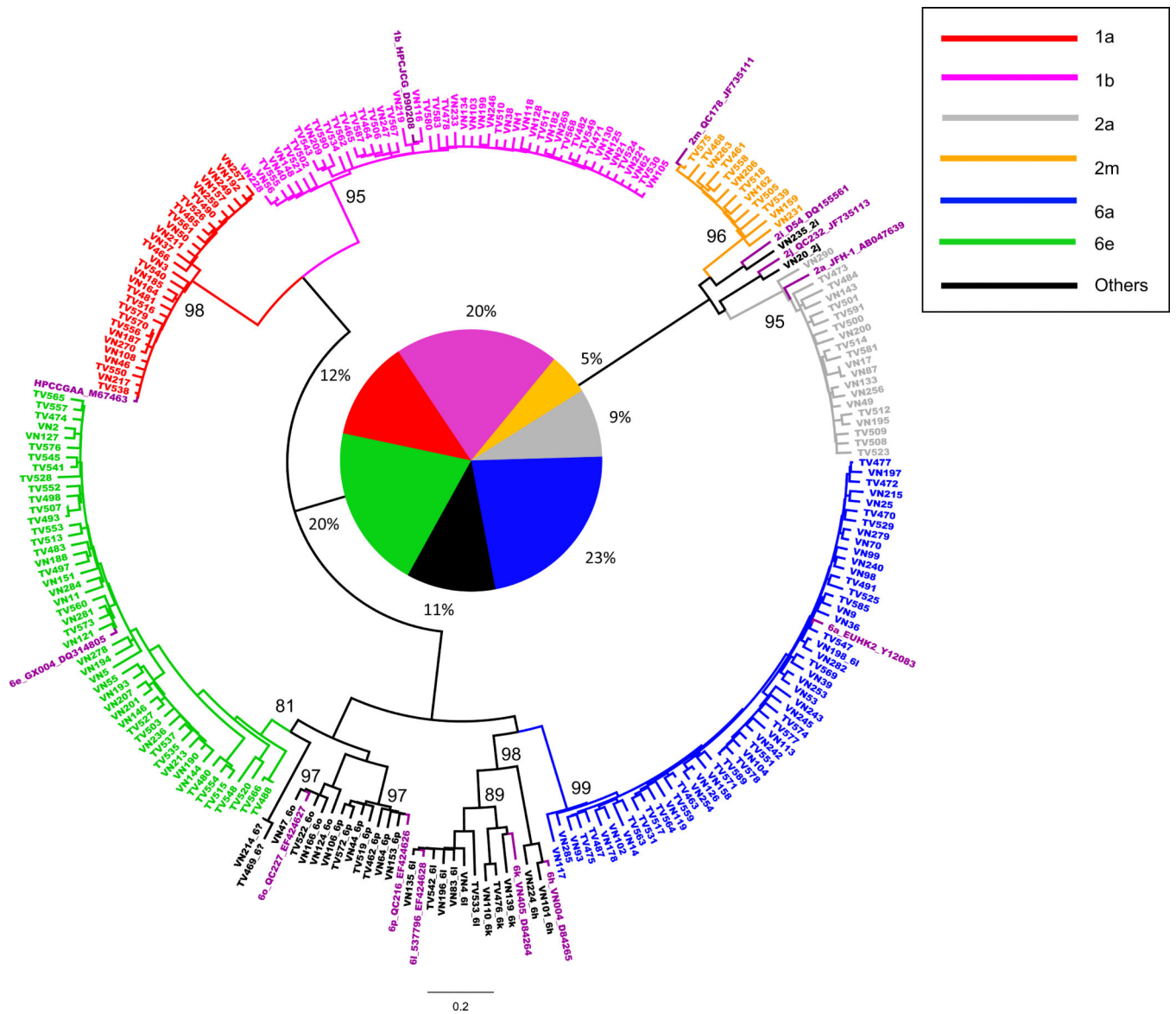
Figure 1.
A circular form of phylogenetic tree based on the Core-E1 sequences from 236 study subjects co-analyzed with 13 references for 13 assigned subtypes. Different subtypes are shown in different colors and indicated on the upper right corner of the tree, while clusters in black show minor subtypes with their labels post fixed. All references are shown in purple with their format as: subtype_isolate ID_GenBank number. Close to the internal nodes of each cluster, two digit numbers show the values of bootstrap analysis, of which only those 81% are presented. In the center of the tree, a pie chart with its neighboring percentage shows the proportions of the different subtypes of the 236 isolates.

Figure 2.
A circle form of phylogenetic tree based on the NS5B sequences from 236 study subjects co-analyzed with 13 references for 13 assigned subtypes. Except that no significant bootstrap supports were shown and the central pie chart was surrounded with the number of sequences each lineage possesses, all other indications are the same as described in Figure 1.

Figure 3.
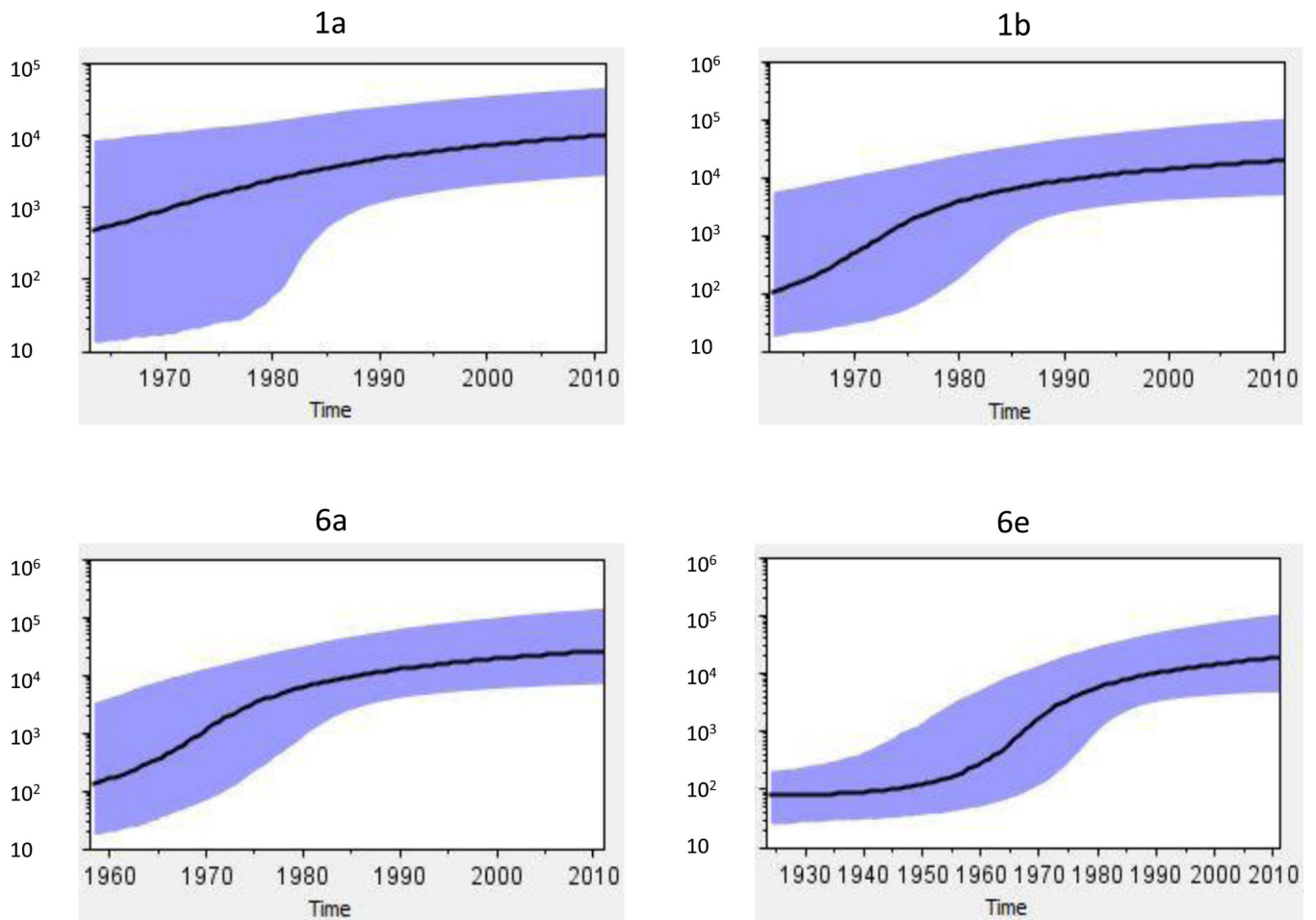Bayesian Skyline Plots. Four diagrams represent four subtypes 1a, 1b, 6a, 6e. The solid line represents the effective population size through time. The purple area represents the 95% highest posterior density confidence interval. The vertical ruler on the left (Y axis) measures the effective population size while the horizontal scale on the bottom measures time from the present (right) to the past (left).
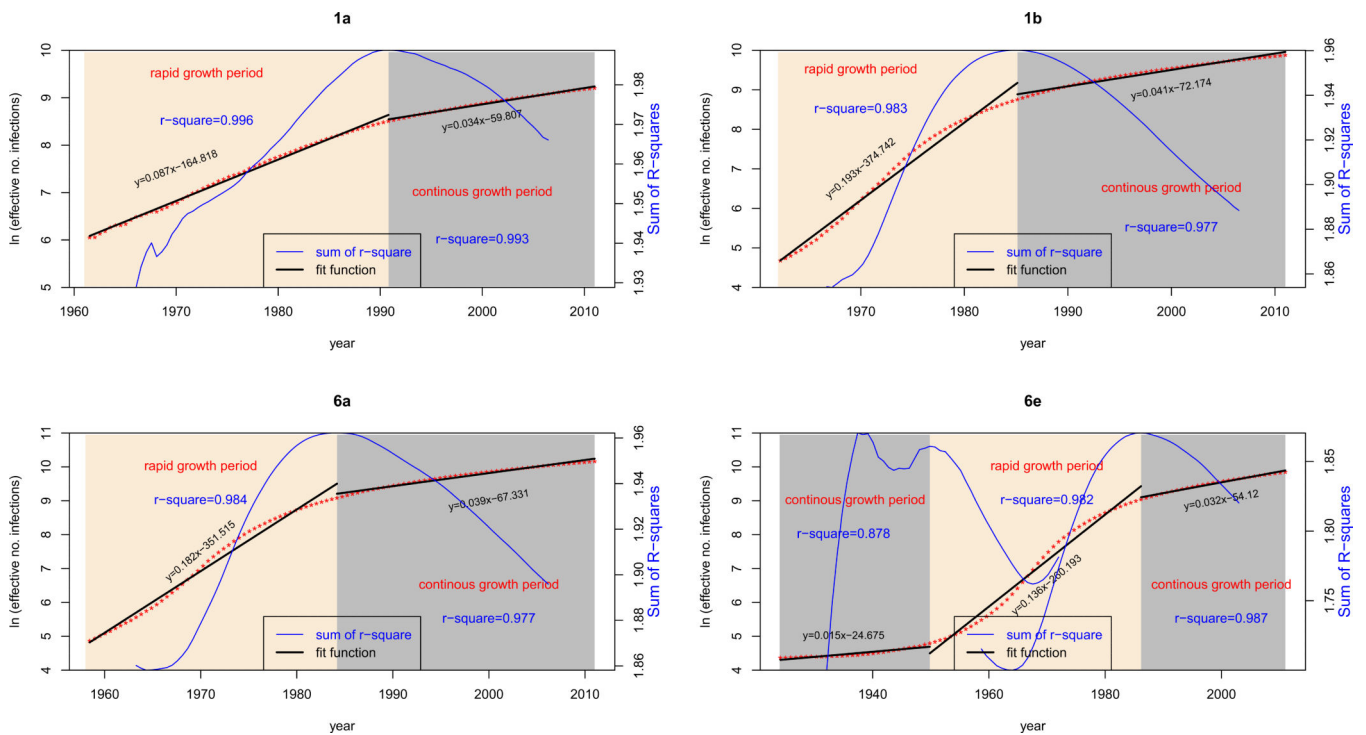
Figure 4.
HCV growth rates. Four simplified diagrams correspond to four BSPs. In each diagram, the median growth curve is shown in red, in which a dot represents a sampling time point. This curve is measured by the vertical ruler for population size on the left and the horizontal time scale on the bottom. The time course of HCV growth is divided into two periods: rapid growth (yellow area) and continuous growth (grey area). For different subtypes, the delimitation of these two periods is dissimilar. Based on the median population sizes exported from the four BSPs within these two periods, simple linear regression analyses were performed and the regression lines are shown separately and appear in each diagram with two discontinued black lines. Accordingly, two regression functions and two R squares are provided in each diagram, in which the slopes indicate the growth rates. In each diagram, a blue curve is also presented, reflecting the changes in the sum of R-squares generated after sliding two regression analyses through all sampling time points within a certain range. Measured by the vertical ruler on the right, this curve identifies the best time point at which the sum of R squares maximizes and two growth periods are thereby divided. Because the BSP for 6e has an additional period of constant population size prior to its rapid growth, three black and two blue curves are shown.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

General information of the 236 study subjects and the detected HCV genotypes

Genotype	1a	1b	2a	2i	2j	2m	6a	6e	eh	6k	6l	6o	6p	6 uc	Total
Total	29	48	20	1	1	12	53	48	2	3	6	4	7	2	236
%	12.29	20.34	8.47	0.42	0.42	5.08	22.46	20.34	0.85	1.27	2.54	1.69	2.97	0.85	100
Male	21	27	12	1	1	4	33	34	1	2	5	2	3	1	146
Female	7	21	8			8	18	14	1	1	1	2	3	1	86
Sex unclear	1						2						1		4
Mean Age	46.7±11.97	46.0±11.43	49.1±10.39	63	48	46.8±12.84	43.9±10.53	46.5±11.13	37.5±12.02	44.3±20.23	45.2±12.83	52±24.67	45.17±10.65	42.5±2.12	46.13±11.49

Table 2

Statistics generated from Bayesian MCMC analyses for estimating the four BSPs

BSP	1a[a] (n=29)	1b[b] (n=48)	6a[a] (n=53)	6e[a] (n=48)
Rate Prior[c]	N (9.1025E-4, 1E-5)	N (1.018E-3,1E-5)	N (1.1632E-3,1E-5)	N (1.1646E-3,1E-5)
Posterior[d]	−3228.62±0.21 (−3255.38, −3202.82)	−5135.33±0.31 (−5166.31, −5105.38)	−6827.13±0.42 (−6858.84, −6794.64)	−6413.74±0.31 (−6441.25, −6385.22)
Alpha	1.4496±2.22E-2 (0.27, 2.78)	0.7083±7.36E-3 (0.33, 1.14)	0.9082±3.23E-3 (0.6, 1.23)	1.0937±3.51E-3 (0.7, 1.47)
pInv	0.5±3.65E-3 (0.33, 0.62)	0.4±2.63E-3 (0.25, 0.51)	0.44±7.48E-4 (0.37, 0.51)	0.45±6.51E-4 (0.38, 0.51)
Uced.mean	9.10E-04±1.16E-07 (8.90E-04, 9.29E-04)	1.02E-03±1.04E-07 (9.99E-04, 1.04E-03)	1.16E-03±1.10E-07 (1.14E-03, 1.18E-03)	1.16E-03±1.04E-07 (1.15E-03, 1.18E-03)
COV	0.72±1.44E-03 (0.58, 0.88)	0.73±1.39E-03 (0.6, 0.87)	0.70±1.15E-03 (0.59, 0.83)	0.70±1.49E-03 (0.57, 0.85)
tMRCAs	83.78±0.53 (48.12, 129.67)	79.97±0.4 (49.27, 115.79)	78.11±0.43 (53.1, 110.24)	168.28±1.35 (87.76, 294.04)
Covariance	−0.04±1.37E-03 (−0.29, 0.21)	−0.02±1.17E-03 (−0.22, 0.18)	−0.08±9.74E-04 (−0.26, 0.09)	−0.07±1.16E-03 (−0.25, 0.12)
Treelikelihood	−3069.34±0.08 (−3082.31, −3055.9)	−4827.15±0.16 (−4845.3, −4809.77)	−6446.43±0.23 (−6466.04, −6427.12)	−6062.87±0.18 (−6081.18, −6045.46)

[a] a normal rate estimated for the actual dataset in this study.

[b] a normal rate estimated in this study for an alternative 1b dataset.

[c] Normal distribution (mean, lower and upper boundaries).

[d] Mean ± Standard error of mean (upper and lower 95% confidence intervals)

Table 3

Statistics related to the estimates of growth rates

Subtype (prior molecular rate)	BSP Regression Slope					conLog.growthRate[d]				conLog.t$_{50}$
	Slope 1[a]	Intercept 1[a]	Switch point[b]	Slope 2[c]	Intercept 2[c]	Mean	95%lower	95%upper	Median	
1a (9.1025E-4)	0.087	−166.658	1990.81	0.034	−60.4199	0.2295	0.1351	0.343	0.2232	34.335
1b (1.018E-3)	0.193	−428.18	1985.12	0.041	−63.23	0.3581	0.212	0.5203	0.3499	34.125
6a (1.1632E-3)	0.182	−381.24	1984.18	0.039	−57.11	0.3001	0.184	0.4233	0.2929	33.916
6e (1.1646E-3)	0.136	−283.02	1986.18	0.0292	−48.77	0.2494	0.1386	0.3702	0.2416	34.036

[a] Estimated for the period of rapid exponential growth.

[b] The best-fitting time point at which the sum of r^2 of two regression analyses maximized, which divides the periods of rapid growth and continuous growth.

[c] Estimated for the period of continuous growth from 1990-the present.

[d] Estimated under the constant logistic model using BEAST.