

Viral Gene Sequences Reveal the Variable History of Hepatitis C Virus Infection among Countries

Tatsunori Nakano,¹ Ling Lu,^{2a} Pengbo Liu,³ and Oliver G. Pybus⁴

¹Department of Internal Medicine, Ichinomiya Nishi Hospital, Ichinomiya, Aichi, Japan; ²Division of Digestive Disease, Department of Medicine, and ³Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia; ⁴Department of Zoology, University of Oxford, Oxford, United Kingdom

Background. The analysis of molecular phylogenies estimated from the gene sequences of sampled viruses can provide important insights into epidemiological processes.

Methods. The demographic and migration histories of the prevalent hepatitis C virus (HCV) subtypes 1a and 1b were inferred from viral gene sequences sampled in 5 countries. Estimated viral phylogenies were analyzed by use of methods based on parsimony and coalescent theory.

Results. The parsimony migration analysis suggested that the global subtype 1a and 1b epidemics are geographically structured, with asymmetrical movement of HCV strains among the sampled countries. The coalescent analysis indicated that subtype 1a infections in the United States, Brazil, and Indonesia began to increase exponentially during the 1940s and 1950s, whereas in Vietnam the increase began after the 1970s. In contrast, subtype 1b infections in these 4 countries and in Japan began to increase exponentially between 1880 and 1920, with a possible recent decrease in infection rates in Indonesia and Japan. In the United States, Brazil, and Vietnam, the epidemic growth rates for subtype 1a strains were higher than those for subtype 1b strains, whereas the growth rates were similar in Indonesia.

Conclusions. The estimated histories of migration and population growth indicated that patterns of HCV transmission differ among countries and viral subtypes.

Hepatitis C virus (HCV) infects >170 million people worldwide [1], the majority of whom are chronically infected. Chronically infected individuals serve as a source of transmission to others and are at risk for severe HCV-related diseases, such as liver cirrhosis and hepatocellular carcinoma, during the decades after initial infection [2]. Therefore, it is important to estimate the demographic history of HCV infection, to predict the future burden of disease.

HCV is classified into 6 major genotypes [3, 4]. Within the genotypes there are many subtypes, with varying geographic distributions and modes of trans-

mission [3]. For example, subtypes 1a and 1b are prevalent worldwide. In Europe, subtype 1a is generally found in younger individuals, who often have injection drug (ID) use as a risk factor, whereas subtype 1b is common in older individuals with a history of blood transfusion [5–7]. In the United States, a statistically significant association between subtype 1a infection and ID use has been reported [8].

Coalescent theory can be used to infer the past epidemic growth of HCV infection [9, 10]. This method estimates the history of changing viral population size, using phylogenetic trees that are reconstructed from sampled viral gene sequences. In a previous study [9], the spread of subtype 1a and 1b infections worldwide was analyzed by use of HCV sequences collated from sequence databases. In the present study, new sequence data were obtained from 5 countries and were used to infer the epidemic and migration history of HCV subtypes 1a and 1b. First, we investigated the degree of phylogenetic clustering of HCV strains by country. Second, we estimated the epidemic histories of HCV in these 5 countries using coalescent theory.

Received 7 November 2003; accepted 23 February 2004; electronically published 10 August 2004.

Financial support: Wellcome Trust (to O.G.P.).

^a Present affiliation: Division of Gastroenterology/Hepatology, Department of Medicine, Kansas University Medical Center, Kansas City.

Reprints or correspondence: Dr. Tatsunori Nakano, Dept. of Internal Medicine, Ichinomiya Nishi Hospital, Okucho Orikuchinishi 89-1, Ichinomiya, Aichi 491-0201, Japan (nakano@anzu.or.jp).

The Journal of Infectious Diseases 2004;190:1098–108

© 2004 by the Infectious Diseases Society of America. All rights reserved.
0022-1899/2004/19006-0010\$15.00

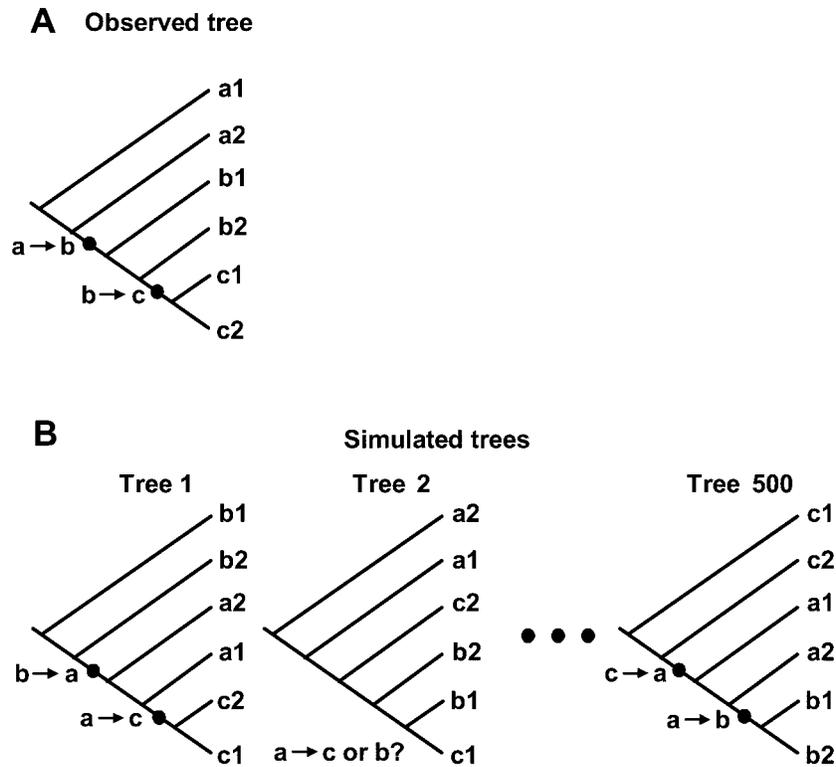


Figure 1. Illustration of the parsimony method used to obtain observed and expected no. of changes in location among countries. The observed tree (A) contains 6 sequences sampled from 3 countries (*a*, *b*, and *c*). The parsimony algorithm identifies 2 changes in location: from *a* to *b* and from *b* to *c*. Using the observed tree topology, the sequence locations are randomized 500 times (B). For each of the randomized trees, changes in location are analyzed by use of the same parsimony algorithm. Differences between the observed and simulated no. of changes represent genetic isolation or migration of hepatitis C virus strains.

MATERIALS AND METHODS

HCV gene sequences from the United States, Brazil, Vietnam, and Indonesia. Originally, 99, 80, 257, and 262 anti-HCV-positive serum samples were randomly selected from the following blood banks, respectively: the New York Blood Center (United States), the Sao Paulo Blood Center (Brazil), the Hanoi Blood Center (Vietnam), and the Jakarta Blood Center (Indonesia). The samples were collected during 1999–2002. Informed consent was obtained from all participants at each blood center. The human-subject guidelines of the Centers for Disease Control and Prevention were strictly followed.

A partial region of the HCV NS5B gene was amplified by nested reverse-transcription polymerase chain reaction, with the following 2 primer pairs: for the first round, P1 (5'-TGGG-GATCCCGTATGATACCCGCTGCTTTG-3' [nt 8245–8275; nucleotide numbering is relative to strain M62321]) and P2 (5'-GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' [nt 8645–8616]); for the second round, P3 (5'-CTCAACCGTCAC-TGAGAGAGACAT-3' [nt 8276–8299]) and P4 (5'-GCTCTCA-GGCTCGCCGCTCCTC-3' [nt 8615–8592]). Cycle sequencing was performed with dRhodamine terminators, and the

products were electrophoresed by use of a DNA sequencer (377A; Applied Biosystems). The gene region obtained has been used extensively for HCV genotyping and molecular epidemiology [3, 9, 11]. Subtyping of viral sequences was performed by phylogenetic analysis with a broad panel of reference sequences. The numbers of subtype 1a and 1b sequences obtained were as follows: 24 subtype 1a and 17 subtype 1b sequences in the United States; 18 subtype 1a and 26 subtype 1b sequences in Brazil; 27 subtype 1a and 32 subtype 1b sequences in Vietnam; and 30 subtype 1a and 37 subtype 1b sequences in Indonesia. The GenBank accession numbers of the sequences obtained in the present study are AY224780–AY224990.

HCV gene sequences from Japan. Japanese subtype 1a and 1b strain sequences were obtained from the Hepatitis Virus Database (<http://s2as02.genes.nig.ac.jp>); however, insufficient subtype 1a sequences were found for further analysis. Full genome sequences isolated from Japanese individuals were selected. Information from the database, from primary literature, and from original authors was used to exclude both multiple sequences from the same individual or from related individuals and sequences that had not been directly isolated from infected

Table 1. Isolation or migration of hepatitis C virus subtype 1a among 4 countries.

Category, country of origin	Destination country			
	United States	Brazil	Vietnam	Indonesia
No. of observed changes in tree (total, 9)				
United States	...	4	2	0
Brazil	0	...	0	0
Vietnam	1	0	...	0
Indonesia	2	0	0	...
No. of expected changes per tree ^a (total, 22.17)				
United States	...	1.31	1.63	1.58
Brazil	0.68	...	0.69	0.77
Vietnam	2.17	1.94	...	2.42
Indonesia	3.12	2.58	3.28	...
Difference between the above observed and expected no. of changes (total difference, -13.17)				
United States	...	2.69	0.37	-1.58
Brazil	-0.68	...	-0.69	-0.77
Vietnam	-1.17	-1.94	...	-2.42
Indonesia	-1.12	-2.58	-3.28	...

NOTE. None of the differences was significant, except for the total of -13.17 ($P < .05$).

^a Calculated on the basis of simulated trees.

individuals. The sampling dates of the serum samples were also obtained, by communication with original authors. Finally, 48 subtype 1b sequences, derived from serum collected during 1993–1997, were selected (accession nos. AB049088–AB049101, AF165045, AF165047, AF165049, AF165051, AF165053, AF165055, AF165057, AF165059, AF165061, AF165063, AF207752–AF207774, and AF208024). For these sequences, partial E1 gene sequences (nt 615–914) were analyzed in addition to the NS5B region.

Analysis of isolation and migration of HCV sequences among countries. First, phylogenies of subtype 1a and 1b sequences from all of the countries under investigation were estimated under maximum likelihood (ML), by use of PAUP software (version 4.0b; Sinauer Associates). The substitution model used was the general time-reversible model, with γ -rate heterogeneity and without a molecular clock. The trees were rooted with other subtypes as an outgroup. Each HCV sequence was labeled according to its country of origin. Second, the migration histories of subtype 1a and 1b infections were inferred from these phylogenies by use of a method based on parsimony theory. Given a phylogeny in which each tip has been designated to be a given “state,” the parsimony algorithm estimates the minimum number of “state changes” needed to give rise to the observed distribution of states (see [12] for a description of the parsimony method). In the present study, the states are the locations of each sequence, and, therefore, state changes represent migration events (see figure 1A for an illustration of the method). Ambiguous changes were not counted. The sampling locations of all sequences were then randomized 500 times, and for each randomization the parsimony number of changes in location was again calculated, in the same manner (figure 1B). The total number of changes was summed across

all 500 simulated trees and then was divided by the number of replicates, which gave the expected number of changes under the null hypothesis of complete geographic mixing of sequences among countries. The difference between the observed and expected number of changes for each pair of countries indicates the level of genetic isolation or migration. The parsimony calculations were performed by use of MacClade software (version 4; Sinauer Associates). Figure 2 shows the phylogenies used in the migration analysis.

Analysis of the epidemic history of HCV by use of coalescent theory. The effective number of infections through time, N_e , was estimated by use of a statistical framework based on coalescent theory [13]. The analysis was performed in 3 steps: (1) HCV phylogenies for each country were estimated (the ML method in PAUP software [version 4.0b] was used, with the HKY85 substitution model, a codon-position model of rate heterogeneity, and the molecular clock enforced); (2) a nonparametric estimate of N_e was calculated from each phylogeny; and (3) parametric ML estimates of N_e were obtained by use of an appropriate demographic model. Steps 2 and 3 were conducted by use of GENIE software (version 3.1) [14]. This program treats time t as zero at present, so that t increases into the past; therefore, N_0 is the effective number of infections at present. Nonparametric estimates of N_e , called “skyline plots,” are obtained by transforming the coalescent times of an observed genealogy into a piecewise plot of effective population size through time [13, 15]. The skyline plot was used as a model-selection tool; it suggested which parametric model was likely to best fit the data (see below). Parametric estimates of effective population size for each phylogeny were obtained by use of 9 different demographic models [14]. Each model had

Table 2. Isolation or migration of hepatitis C virus subtype 1b among 5 countries.

Category, country of origin	Destination country				
	United States	Brazil	Vietnam	Indonesia	Japan
No. of observed changes in tree (total, 20)					
United States	...	0	0	0	2
Brazil	0	...	0	1	0
Vietnam	2	1	...	2	0
Indonesia	1	2	1	...	2
Japan	1	3	0	2	...
No. of expected changes per tree ^a (total, 38.08)					
United States	...	0.25	0.28	0.25	0.24
Brazil	0.57	...	0.84	0.90	0.87
Vietnam	1.01	1.41	...	1.69	1.78
Indonesia	1.37	2.10	2.33	...	2.55
Japan	3.44	4.85	5.44	5.91	...
Difference between the above observed and expected no. of changes (total difference, -18.08)					
United States	...	-0.25	-0.28	-0.25	1.76
Brazil	-0.57	...	-0.84	0.10	-0.87
Vietnam	0.99	-0.41	...	0.31	-1.78
Indonesia	-0.37	-0.10	-1.33	...	-0.55
Japan	-2.44	-1.85	-5.44	-3.91	...

NOTE. The differences for the United States to Japan (1.76), for Japan to Vietnam (-5.44), and for the total (-18.08) were statistically significant ($P < .05$); all other differences were not significant.

^a Calculated on the basis of simulated trees.

1–4 demographic parameters. Likelihood ratio tests were used to test the goodness-of-fit of the parametric models. The ML estimates and 95% confidence intervals (CIs) of demographic parameters were estimated by use of the ML framework implemented in GENIE software (version 3.1). The estimates of past population size were transformed into a calendar time scale based on years by use of the substitution rate of the viral gene concerned. For subtype 1b, we used substitution rates that had been previously estimated: 5.0×10^{-4} substitutions/site/year for the NS5B region and 7.9×10^{-4} substitutions/site/year for the E1 region [9].

RESULTS

Analysis of isolation and migration of HCV sequences among countries. Figure 2 shows the phylogenies of the HCV subtype 1a (figure 2A) and 1b (figure 2B) sequences obtained in the present study. As can be seen in figure 2A, clusters of Vietnamese and Brazilian sequences were found. In addition to these clusters, 4 other clusters were found, which mainly contained sequences from Indonesia and the United States, the United States and Brazil, Indonesia and Vietnam, and the United States and Brazil. As can be seen in figure 2B, clusters that were mainly composed of HCV sequences from a single country were found. In summary, both subtypes show grouping of sequences by country, but each country tends to contain >1 HCV lineage. This finding indicates intermediate levels of phylogenetic mixing among countries.

To measure country-wise clustering statistically, the isolation and migration of subtype 1a and 1b sequences were analyzed by use of a parsimony method. The estimated number of changes in location between countries (i.e., migration events) was 9 for subtype 1a and 20 for subtype 1b, whereas the expected number of location changes for the 500 simulated trees created with randomized locations was 22.17 for subtype 1a and 38.08 for subtype 1b (tables 1 and 2). The observed number of migration events was significantly smaller ($P < .05$) than that expected under the null hypothesis of complete geographic mixing; therefore, this hypothesis can be rejected. This result suggests that there is considerable subdivision by location among the subtype 1a and 1b strains sampled.

The parsimony analysis also provided clues about the movement of subtype 1a and 1b strains among the 5 countries sampled. Tables 1 and 2 show the difference between the expected and observed number of changes for each pair of countries. Negative values indicate less movement from one country to another than that expected by chance, whereas positive values indicate more movement than expected. For subtype 1a, most of the differences were negative, suggesting little movement of HCV strains among the countries. In particular, HCV lineages sampled in Vietnam and Indonesia showed little movement to other countries. There was a tendency toward more movement of HCV lineages from the United States to Brazil than was expected ($P = .08$), but movement from Brazil to the United States was not more common than expected. For subtype 1b,

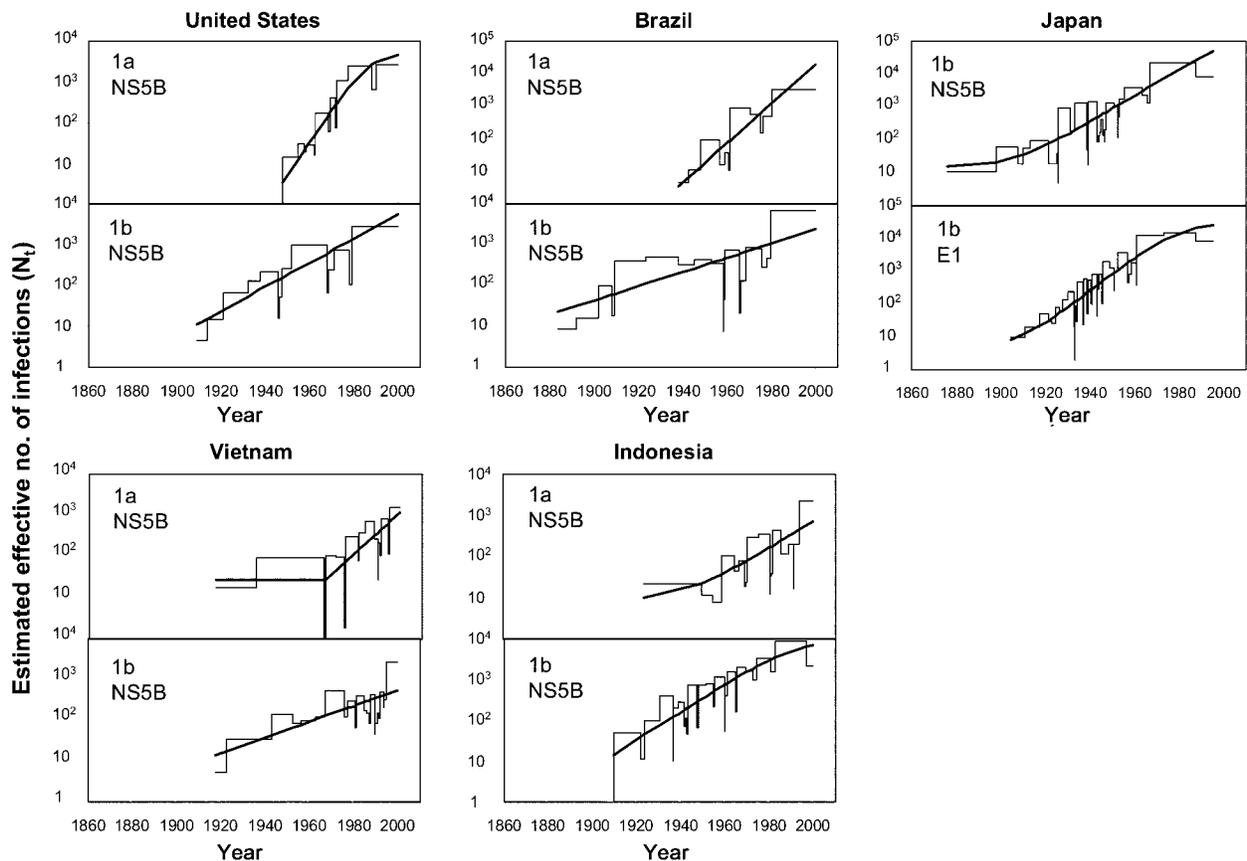


Figure 3. Effective no. of hepatitis C virus (HCV) infections through time (N_t), as estimated from the phylogenies of each HCV population. The thin, stepwise plots are nonparametric estimates of N_t , called “skyline plots.” The bold lines show parametric estimates of N_t , obtained by use of an appropriate demographic model. The estimates were transformed into a time scale based on years by use of estimated rate of evolution (see text).

many of the values were negative, again suggesting less overall migration than would be expected by chance. The Japanese 1b strains contain “imported” lineages, as indicated by the positive value for movement from the United States to Japan ($P < .05$).

Analysis of the epidemic history of HCV by use of coalescent theory. The level of population subdivision demonstrated in the parsimony analysis above suggests that much of the transmission of the sampled HCV strains occurred within the sampled countries. Therefore, the epidemic history of subtypes 1a and 1b in each country was estimated from separate trees, each of which contained sequences from only 1 country.

Table 3 shows, for each HCV population, the best-fit demographic model and the ML estimates (with 95% CIs) for each demographic parameter. Different demographic models fit different populations. Figure 3 shows these estimates graphically and compares them with nonparametric estimates of N_t (skyline plots). The common ancestor of United States 1a and Brazil 1a appeared around 1940–1950, and these lineages show rapid exponential growth, although the growth rate of United States 1a has decreased since 1990. Vietnam 1a strains originated around 1920, and there is no evidence indicating ex-

ponential growth until around 1970, after which they grew exponentially until the present. Indonesia 1a strains originated around 1920 and show an increasing rate of exponential growth through time. In contrast, subtype 1b strains in all 5 countries have common ancestors that are considerably older, dating to 1880–1920. All countries show an exponential increase in the number of subtype 1b infections, with a decrease in growth in the recent past for Indonesia and Japan. The epidemic histories estimated from the NS5B and E1 regions of the same Japan 1b strains show similar rates of growth, but the E1 data contain a stronger signal for a recent slowdown in infection.

The exponential growth rates varied by populations and subtypes (table 3). The estimated rates for United States 1a and Brazil 1a are the highest, followed by those for Vietnam 1a and Japan 1b; the rates for Indonesia 1a and 1b and United States 1b are slightly lower. The Vietnam 1b and Brazil 1b epidemics grew at the lowest rates. For the United States, Brazil, and Vietnam, the rates for subtype 1a are significantly higher than those for subtype 1b.

The contribution of subtypes 1a and 1b to HCV prevalence in different countries. Current estimates of the HCV subtype

Table 3. Analyzed genome region, no. of sequences, selected demographic models, equations, and maximum-likelihood (ML) estimates and 95% confidence intervals (CIs) of each demographic parameter for each hepatitis C virus population.

Parameter	United States		Brazil	
	1a	1b	1a	1b
Genome region	NS5B	NS5B	NS5B	NS5B
No. of sequences	24	17	18	26
Model	LG	EG	EG	EG
Equation	$N_t = N_0 \{ [1 + c] / [1 + c \cdot \exp(rt)] \}$	$N_t = N_0 \cdot \exp(-rt)$	$N_t = N_0 \cdot \exp(-rt)$	$N_t = N_0 \cdot \exp(-rt)$
N_0	4312 (990–120,868)	5505 (1536–26,943)	18,834 (3948–120,416)	2375 (1123–5762)
r	0.185 (0.129–0.244)	0.068 (0.045–0.093)	0.140 (0.100–0.182)	0.040 (0.027–0.053)
α
β
c	0.092 (0.0030– ∞)

NOTE. Data are ML estimates (95% CIs), unless otherwise noted. α , Population size at $t = \infty$ as a proportion of N_0 ; β , population size before change as a proportion of N_0 ; c , logistic shape parameter; EG, exponential growth; EXG, expansion growth; GLG, generalized logistic growth; LG, logistic growth; N_0 , population size at present; PEXG, piecewise expansion growth; r , exponential growth rate; t , time.

^a Where $Z_t = \{ [1 + c] / [1 + c \cdot \exp(rt)] \}$.

distribution in each sampled country [16–19] were used to transform the epidemic histories shown in figure 3 to reflect the relative historical levels of subtype 1a and 1b infection (figure 4). If the current proportion of a particular subtype in a particular country is $y\%$, then the historical level of infection by that subtype at time t in the past is $(N_t \times y) / N_0$, where N_t and N_0 are the effective numbers of infections at time t and time 0, respectively.

As figure 4 shows, subtype 1b was predominant in the United States before the 1970s, after which subtype 1a became more frequent, although the growth rate of subtype 1a appears to have slowed recently. In Brazil, subtype 1b is predominant, but subtype 1a infections have increased in relative frequency in recent years. In Vietnam, subtype 1b had been predominant for many years, but again subtype 1a shows a recent increase in frequency, after 1990. In Indonesia, subtype 1b is predominant, and the dramatic growth of subtype 1a observed in the other countries is not seen.

DISCUSSION

The analysis of HCV subtype 1a and 1b strains presented here provides useful information about the epidemiology of HCV, both in space and across time. Analysis of isolation and migration of the sampled strains has shown country-wise clustering and some movement of the strains among different countries. Investigation of the demographic histories of subtypes 1a and 1b has demonstrated variation in the pattern of HCV spread among countries.

The international migration of HCV strains is caused by the movement of infected individuals through travel and immigration and by the import and export of contaminated blood products. Cochrane et al. recently reported subtype 1a transmission among countries through ID use [20]. In the present

study, the observed 1-way movement of subtype 1a from the United States to Brazil, not seen for subtype 1b, may represent transmission by ID users. Vietnam 1a and Indonesia 1a strains originated before United States 1a and Brazil 1a strains, yet movement of Vietnamese or Indonesian strains to the United States or Brazil was less common than expected. This pattern of movement makes it unlikely that the subtype 1a epidemics in Western countries were derived from Asian countries. The movement of subtype 1b from the United States to Japan is suggested by our results; therefore, some HCV strains in Japan seem to have been introduced from the United States, although our analysis did not evaluate the time of introduction. HCV infections in Japan caused by imported blood products from the United States have been reported [21]. Further analysis using more sequences from more countries will provide more-reliable and more-detailed information about the movement of HCV strains.

In a previous analysis, no constant association between phylogeny and the geographic origin of subtype 1b sequences was reported [11]. For subtype 1a and 1b strains, the present analysis has shown intermediate levels of population subdivision among countries. The data suggest that much of the transmission history represented by the phylogenies shown in figure 2 occurred within individual countries, although each country tended to contain >1 HCV lineage. As a result, we should interpret the results of the coalescent analysis carefully; some of the divergence near the root of the phylogeny for a given country is likely to represent transmission that occurred outside that country. This is particularly true for the periods of slow (or zero) population growth observed for Vietnam 1a, Indonesia 1a, and Japan 1b. Therefore, when one interprets population growth within each country, it is more informative to focus on the exponential growth phase and afterward, rather

Vietnam		Indonesia		Japan	
1a	1b	1a	1b	1a	1b
NS5B	NS5B	NS5B	NS5B	NS5B	E1
27	32	30	37	48	48
PEXG	EG	EXG	LG	GLG	GLG
$N_i = N_0 \cdot \exp(-rt)$ if $t < -\ln(\beta)/r$; otherwise, $N_i = N_0\beta$	$N_i =$ $N_0 \cdot \exp(-rt)$	$N_i =$ $N_0\{\alpha + [(1 - \alpha) \cdot \exp(-rt)]\}$	$N_i =$ $N_0\{[1 + c]/[1 + c \cdot \exp(rt)]\}$	$N_i =$ $N_0\{\alpha + [(1 - \alpha)Z_i]^p\}$	$N_i =$ $N_0\{\alpha + [(1 - \alpha)Z_i]^p\}$
1180 (533–3847)	546 (323–990)	750 (370–1793)	6767 (1863–35,050)	48,157 (6520–245,280)	26,885 (5789–479,298)
0.112 (0.073–0.186)	0.045 (0.027–0.061)	0.078 (0.044–0.145)	0.080 (0.059–0.112)	0.094 (0.070–0.135)	0.119 (0.091–0.178)
...	...	0.011 (0–0.080)	...	0.00031 (0–0.0049)	0.00017 (1.1 × 10 ⁻³⁰ –0.00037)
0.22 (0.0085–∞)
...	0.51 (0.014–∞)	4.6 (0.022–∞)	0.17 (0.0025–∞)

than to emphasize the dates of origin. If larger data sets become available in the future, then this complication could be avoided by the estimation of separate epidemic histories for each lineage (or cluster) from each country. In addition, this approach could provide estimates of the introduction time for each lineage in each country.

In our coalescent analysis, we applied substitution rates previously estimated for subtype 1b to both subtypes 1a and 1b [9]. A very similar substitution rate for subtype 1a, estimated by use of similar methodology, was also recently reported [22]. Many investigators have studied the molecular epidemiology of HCV using the same genome regions as those used here, with similar substitution rates, and have shown that their results are consistent with what is known about the epidemic history of HCV [3, 9–11, 22, 23]. There is no evidence to suggest that HCV substitution rates vary significantly among genotypes or subtypes. This is also true for HIV: the origins of HIV-1 and HIV-2 have been successfully inferred on the assumption of a single substitution rate among subtypes [24, 25]. However, substitution rates do vary strongly among HCV genes—a result of varying selective constraint—which is why, in the present study, we were careful to use substitution rates that matched the genome regions under investigation. We note that there is uncertainty in our substitution-rate estimates (see [9] for details) and that this uncertainty is not incorporated into the coalescent results (table 3 and figure 3). However, this uncertainty is relatively small and, thus, does not qualitatively affect our results.

In Japan, subtype 1b is most prevalent and accounts for 70% of all HCV infections [26]. Initial HCV spread among ID users in the 1950s, unsafe injection practices in medical settings, and contaminated blood transfusions from the 1960s to the early 1970s are thought to be the main transmission routes of HCV [27]. New chronic HCV infections are now very rare, a result of the abrogation of paid blood donation in 1968 and the screening of donated blood for hepatitis B surface antigen and anti-HCV beginning in 1973 and 1989, respectively [27, 28]. Japanese subtype 1b E1 gene sequences indicate a decrease in

HCV transmission after 1970. Our estimates based on the E1 and NS5B regions show the exponential growth of subtype 1b in Japan since 1910–1920. Before the widespread use of blood transfusion, several kinds of medical interventions were frequently followed by cases of hepatitis. In Japan, for example, it is known that since 1923 schistosomiasis has been widely treated with intravenous antimony sodium tartrate injection [29]. HCV infection may have started to spread in Japan due to such medical interventions. The corroboration of our estimated population histories with known epidemiological data suggest that the methodology and evolutionary rates we used are reliable.

In the United States, prevalent subtypes in the general population are 1a (52%) and 1b (26%) [16]; an association between subtype 1a infection and ID use has been reported [8]. Our estimates of the epidemic history of 1a in the United States are consistent with the onset of injection opiate use between 1950 and 1960 [30] and with the decrease in newly acquired HCV infections between 1989 and 1995, which has been correlated with a decrease in ID users [31], confirming that ID use is the main transmission route of subtype 1a in the United States. The relative importance of ID use and blood transfusion associated with HCV transmission in the United States has changed over time [32], and ID use has been the predominant mode since the 1970s [33]. This change is consistent with our results (figure 4), which show a change in the predominant subtype from 1b to 1a during the 1970s. Although an association between subtype 1b infection and blood transfusion in the United States has not been reported, some studies have suggested such an association in several European countries [5–7].

Our estimated exponential growth rates may vary according to the transmission route of each strain. The United States 1a epidemic grew at the highest rate. It is known that ID users rapidly acquire HCV infection after initiating injection practice, and up to 90% of them are chronically infected [33]. The intensity and quickness of the acquisition of the infection in ID users are reflected in the high growth rate. Relatively lower

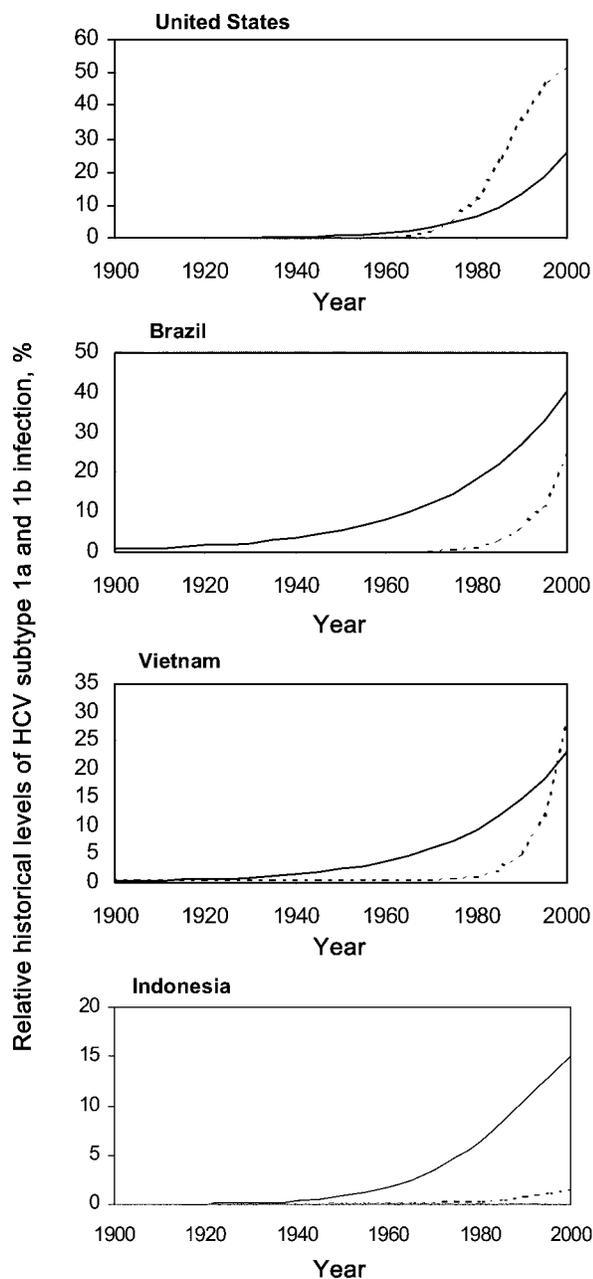


Figure 4. Relative historical levels of hepatitis C virus (HCV) subtype 1a or 1b infection in each country. Estimated levels of subtype 1a infection (*broken lines*) and subtype 1b infection (*solid lines*) were calculated for each country, as described in the text. The curves show the relative levels of subtype 1a and 1b infection at time t in the past, compared with the present proportion of HCV subtypes in each country. Total level of HCV infection for all genotypes at present is 100%.

growth for United States 1b may have resulted from its association with blood transfusion. The growth rate of Japan 1b was between that of United States 1a and that of United States 1b, which is consistent with the information that the Japan 1b epidemic is related to both ID use and blood transfusion.

In Brazil, Vietnam, and Indonesia, the epidemiology of HCV

infection is less well characterized, compared with that in Japan and the United States. In Brazil, the predominant types are subtype 1b (40%), genotype 3 (30%), and subtype 1a (24%) [17], and ID use and blood transfusion have been identified as major transmission modes [34]. The association between transmission route and viral subtype has not been well investigated, although an association between genotype 3 infection and ID use has been reported [35]. In European countries, subtypes 3a and 1a have spread through similar transmission networks of ID users [20]. Our estimated epidemic growth of Brazil 1a (which is similar to that of United States 1a, with an introduction around 1940 and a high growth rate) suggests that there may be an association between 1a infection and ID use in Brazil.

In Vietnam, the anti-HCV-positive rate in blood donors was 20.6% in Ho Chi Minh City and 0.8% in Hanoi in 1992. This large difference in prevalence between southern and northern cities has been attributed to differences in ID use [18]. In Ho Chi Minh City, ID use has been common since the Vietnam War, whereas ID use in Hanoi is thought to have been lower until recently [18]. Our samples were derived from Hanoi and indicate that exponential growth of subtype 1a started after the Vietnam War, with a dramatic increase after 1960. In Vietnam, an appreciable proportion of blood donations have been made by paid donors [36]. ID users are suspected to have donated blood contaminated with HCV, thereby transmitting infection to blood-transfusion recipients [37]. Thus, subtype 1a may have spread in northern Vietnam through both ID use and blood transfusion after the Vietnam War. According to Kakumu et al., subtype 1b is more prevalent than 1a in the general (non-ID use) Vietnamese population [38]. The slow spread of subtype 1b in Vietnam might be mainly dependent on medical intervention and on other unknown transmission routes that cause community-acquired HCV infection.

In Indonesia, modern medicine (surgery, blood transfusion, intravenous medication, and hemodialysis) has been identified as a transmission route of HCV infection [39, 40]. Subtype 1a infection is rare, but an association between this subtype and hemodialysis has been reported [19, 40, 41]. Blood transfusion is a significant cause of HCV infection in patients on hemodialysis [42]. Our estimate of the epidemic history of Indonesia 1a did not show the rapid increase seen for United States 1a, Brazil 1a, or Vietnam 1a. Moreover, the growth rate was similar to that of United States 1b, suggesting that the spread of subtype 1a in Indonesia seems to be related to hemodialysis or blood transfusion rather than to ID use. On the other hand, subtype 1b and 2b infections account for a large number of infections in Indonesia [19, 40, 41]. The growth rate of Indonesia 1b had been similar to that of Indonesia 1a and then decreased in the recent past, possibly as a result of the initiation of donor screen-

ing, which would be consistent with the notion that this epidemic depends mainly on transmission by blood transfusion.

It could be argued that the HCV sequences obtained in the present study are not wholly representative of the sampled countries, due to a small sample size or to risk group-specific sampling. We cannot rule out the possibility that there are unsampled clusters in some countries; however, our coalescent results are consistent with known epidemiological information from the United States and Japan, and the data set presented here is larger than that used in previous studies. A unique advantage of the coalescent approach to molecular epidemiology used here is that the entire history of a transmission cluster can be investigated using a relatively small sample of gene sequences.

Our analysis suggests that subtype 1a, which appears to be associated with ID use, poses the greatest threat to public health in the countries studied. Our results also indicate that blood transfusion has significantly contributed to HCV transmission and that donor screening against HCV transmission has been effective. Hemodialysis, poor infection-control practices, and unsafe injections given in medical settings also have been important in the spread of HCV [33, 43]. Appropriate measures to prevent HCV transmission and to provide medical care are required, not only in the countries studied, but in all other countries with similar conditions.

Acknowledgments

We thank B. H. Robertson and H. A. Fields (Centers for Disease Control and Prevention), for supplying serum samples and for critical manuscript review, and S. Mishiro and K. Nagayama, for providing information about their sequences submitted to GenBank.

References

1. World Health Organization. Global surveillance and control of hepatitis C. *J Viral Hepat* **1999**; 6:35–47.
2. Seeff LB. Natural history of hepatitis C. *Hepatology* **1997**; 26(Suppl 1): S21–8.
3. Simmonds P, Holmes EC, Cha TA, et al. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* **1993**; 74:2391–9.
4. Bukh J, Purcell RH, Miller RH. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc Natl Acad Sci USA* **1994**; 91:8239–43.
5. Silini E, Bono F, Cividini A, et al. Molecular epidemiology of hepatitis C virus infection among intravenous drug users. *J Hepatol* **1995**; 22:691–5.
6. Pawlotsky JM, Tsakiris L, Roudot TF, et al. Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *J Infect Dis* **1995**; 171:1607–10.
7. 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**; 25:1490–6.
8. 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**; 94:3015–9.
9. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science* **2001**; 292: 2323–5.
10. 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**; 20:381–7.
11. Smith DB, Pathirana S, Davidson F, et al. The origin of hepatitis C virus genotypes. *J Gen Virol* **1997**; 78:321–8.
12. Felsenstein J. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates, **2003**.
13. Pybus OG, Rambaut A, Harvey PH. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **2000**; 155:1429–37.
14. Pybus OG, Rambaut A. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* **2002**; 18:1404–5.
15. Strimmer K, Pybus OG. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol Biol Evol* **2001**; 18: 2298–305.
16. Nainan OV, Gao F, Alter M, Margolis H. Determination of the genetic relatedness of HCV isolates: a comparison of the NS5B and HVR1 regions [abstract P-285]. In: Program and abstracts of the 9th International Meeting on HCV and Related Viruses (San Diego, CA). San Diego, CA: University of California, San Diego, **2002**: 220.
17. Campiotto S, Pinho JRR, Carrilho FJ, da Silva LC, Bernardini AP. Hepatitis C virus (HCV) genotypes distribution in different Brazilian regions [abstract V-1014]. In: Program and abstracts of the XIIth International Congress of Virology (Paris). International Union of Microbiological Societies, **2002**:339–40.
18. Song P, Duc DD, Hien B, et al. Markers of hepatitis C and B virus infections among blood donors in Ho Chi Minh City and Hanoi, Vietnam. *Clin Diagn Lab Immunol* **1994**; 1:413–8.
19. Inoue Y, Sulaiman HA, Matsubayashi K, et al. Genotypic analysis of hepatitis C virus in blood donors in Indonesia. *Am J Trop Med Hyg* **2000**; 62:92–8.
20. Cochrane A, Searle B, Hardie A, et al. A genetic analysis of hepatitis C virus transmission between injection drug users. *J Infect Dis* **2002**; 186:1212–21.
21. Kinoshita T, Miyake K, Okamoto H, Mishiro S. Imported hepatitis C virus genotypes in Japanese hemophiliacs. *J Infect Dis* **1993**; 168:249–50.
22. Tanaka Y, Hanada K, Mizokami M, et al. A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. *Proc Natl Acad Sci USA* **2002**; 99:15584–9.
23. Ray SC, Arthur RR, Carella A, Bukh J, Thomas DL. Genetic epidemiology of hepatitis C virus throughout Egypt. *J Infect Dis* **2000**; 182:698–707.
24. Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science* **2000**; 288:1789–96.
25. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci USA* **2003**; 100:6588–92.
26. Okamoto H, Mishiro S. Genetic heterogeneity of hepatitis C virus. *Intervirology* **1994**; 37:68–76.
27. Yoshizawa H. Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. *Oncology* **2002**; 62(Suppl 1):8–17.
28. Nishioka K. Epidemiological studies on hepatitis C virus infection: detection, prevalence, exposure and prevention. *Intervirology* **1994**; 37:58–67.
29. Iida F, Iida R, Kamijo H, et al. Chronic Japanese schistosomiasis and hepatocellular carcinoma: ten years of follow-up in Yamanashi Prefecture, Japan. *Bull World Health Organ* **1999**; 77:573–81.
30. Chambers CD. Some epidemiological considerations of onset of opiate use in the United States. In: Josephson EC, Carroll EE, eds. *Drug use: epidemiological and sociological approaches*. Washington: Hemisphere Publishing, **1973**:65–82.
31. Alter MJ. Epidemiology of hepatitis C. *Hepatology* **1997**; 26(Suppl 1): S62–5.
32. Alter MJ. Hepatitis C virus infection in the United States. *J Hepatol* **1999**; 31(Suppl 1):88–91.
33. Alter MJ, Hutin YJF, Armstrong GL. Epidemiology of hepatitis C virus.

- In: Liang TJ, Hoofnagle JH, eds. Hepatitis C. San Diego: Academic Press, 2000:169–83.
34. Brandao AB, Costa FS. Risk factors for hepatitis C virus infection among blood donors in southern Brazil: a case-control study. *BMC Gastroenterol* 2002; 2:18.
 35. Oliveira ML, Bastos FI, Sabino RR, et al. Distribution of HCV genotypes among different exposure categories in Brazil. *Braz J Med Biol Res* 1999; 32:279–82.
 36. Quan VM, Chung A, Long HT, Dondero TJ. HIV in Vietnam: the evolving epidemic and the prevention response, 1996 through 1999. *J Acquir Immune Defic Syndr* 2000; 25:360–9.
 37. Nakata S, Song P, Duc DD, et al. Hepatitis C and B virus infections in populations at low or high risk in Ho Chi Minh and Hanoi, Vietnam. *J Gastroenterol Hepatol* 1994; 9:416–9.
 38. Kakumu S, Sato K, Morishita T, et al. Prevalence of hepatitis B, hepatitis C, and GB virus C/hepatitis G virus infections in liver disease patients and inhabitants in Ho Chi Minh, Vietnam. *J Med Virol* 1998; 54:243–8.
 39. Sulaiman HA, Julitasari, Sie A, et al. Prevalence of hepatitis B and C viruses in healthy Indonesian blood donors. *Trans R Soc Trop Med Hyg* 1995; 89:167–70.
 40. Hadiwandowo S, Tsuda F, Okamoto H, et al. Hepatitis B virus subtypes and hepatitis C virus genotypes in patients with chronic liver disease or on maintenance hemodialysis in Indonesia. *J Med Virol* 1994; 43:182–6.
 41. Soetjipto, Handajani R, Lusida MI, et al. Differential prevalence of hepatitis C virus subtypes in healthy blood donors, patients on maintenance hemodialysis, and patients with hepatocellular carcinoma in Surabaya, Indonesia. *J Clin Microbiol* 1996; 34:2875–80.
 42. Masuko K, Okuda K, Meguro T, et al. Hepatitis C virus antibodies, viral RNA and genotypes in sera from patients on maintenance haemodialysis. *J Viral Hepat* 1994; 1:65–71.
 43. Madhava V, Burgess C, Drucker E. Epidemiology of chronic hepatitis C virus infection in sub-Saharan Africa. *Lancet Infect Dis* 2002; 2: 293–302.