Population genetic history of hepatitis C virus 1b infection in China

Tatsunori Nakano,† Ling Lu,‡ Yunshao He,§ Yongshui Fu,¶ Betty H. Robertson∥ and Oliver G. Pybus††

†Department of Internal Medicine, Ichinomiya Nishi Hospital, Okucho Origuchinishi 89-1, Ichinomiya, Aichi 491-0201, Japan
‡Division of Gastroenterology/Hepatology, Department of Medicine, Kansas University Medical Center, Kansas City, KS, USA
§Da-An Diagnostic Center, Sun Yat-sen University, Guangzhou, Guangdong, China
¶Guangzhou Blood Center, Guangzhou, Guangdong, China
∥National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA
††Department of Zoology, University of Oxford, Oxford, UK

Subtype 1b is the most common strain of Hepatitis C virus (HCV) in China. Here, the molecular epidemiology and epidemic history of this strain were investigated by conducting phylogenetic and population genetic analyses of E1 and NS5B gene sequences sampled from nine Chinese cities. The phylogenetic analysis indicated the presence of two clusters of Chinese strains that did not include reference strains from other countries, suggesting that these clusters represent two independent chains of HCV transmission within China. The remaining Chinese isolates were more closely related to reference strains from other countries. The date of origin and past population dynamics of the two groups were investigated using a new population genetic method, the Bayesian skyline plot. The estimated dates of origin of both groups coincide with the period of the Chinese ‘Cultural Revolution’ during the years 1966–1976. Both groups grew at a rapid exponential rate between ~1970 and ~1990, after which transmission slowed considerably. Possible explanations for the groups’ fast spread and subsequent slowdown are discussed, including parenteral transmission by unsafe injection, iatrogenic transmission by infected blood or blood products and improvements in blood safety since 1990. These results shed light on HCV transmission in China and may help to predict the future burden of HCV-related disease in the country.

INTRODUCTION

Hepatitis C virus (HCV) is a genetically diverse RNA virus with a single-stranded, positive-sense genome. The virus is classified into six major genotypes with closely related isolates within each genotype being grouped into subtypes (Simmonds et al., 1993; Robertson et al., 1998). Partial genome sequences, particularly those from the E1 and NS5B genes, are commonly used in HCV genotyping and evolutionary analysis (e.g. Simmonds et al., 1993; Smith et al., 1997). Different HCV genotypes and subtypes have different geographical distributions, transmission routes and rates of spread (Bukh et al., 1993; Smith et al., 1997; Pybus et al., 2001). ‘Epidemic’ subtypes, such as 1a, 1b and 3a, are typically found at high prevalences globally having spread rapidly during the twentieth century, probably via infected blood, blood products and injecting drug use. In contrast, ‘endemic’ strains of HCV are usually less prevalent, found in restricted geographical areas and represent long-term, low-level endemic infection in particular populations (Simmonds & Smith, 1997; Pybus et al., 2001).

Evolutionary analysis of sampled virus sequences has advanced considerably during the past decade and is now an important tool in molecular epidemiology. In addition to standard phylogenetic analysis, population genetic methods based on coalescent theory can be used to estimate epidemiological history from viral gene sequences. Coalescent
theory is a stochastic model that describes how population processes, such as changing numbers of infections, determine the shape of viral phylogenies. It therefore enables us to reconstruct the past history of virus transmission from observed virus gene sequences and phylogenies. Coalescent methods have proved particularly useful in reconstructing the history of HCV infection prior to the identification of the virus in 1989 (Pybus et al., 2001, 2003, 2005; Tanaka et al., 2002, 2004, 2005; Nakano et al., 2004).

It has been estimated that 2.5–4.9% of China’s population is HCV-positive (WHO, 2000), although it is possible that some estimates reflect urban populations with higher prevalences than the country as a whole. With an estimated population of 1.3 billion, the number of HCV-infected people in China could thus potentially exceed 50 million. HCV infection is the second most common cause of chronic hepatitis, cirrhosis and hepatocellular carcinoma in China (Tao et al., 1991; Wang et al., 1993; Chen et al., 2002). Recently, the distribution of HCV genotypes in China has been investigated and the predominant strain in the areas studied was subtype 1b (Lu et al., 2005), consistent with earlier studies (Wang et al., 1993; Chen et al., 2002). In addition, Lu et al. (2005) identified two clades of HCV 1b sequences that were prevalent in most regions of China. Here, we analyse Chinese subtype 1b strains together with isolates from other countries. Our phylogenetic analysis suggests the presence of two clusters of HCV (strains of groups A and B), representing two independent chains of HCV transmission in China. Furthermore, we use methods based on coalescent theory to estimate the age of strains from group A and B, and to estimate the historical rates at which these strains spread through the Chinese population. By helping to understand the history of HCV transmission in China, our results may inform HCV control initiatives and can help to infer the future burden of HCV-related disease in the country.

Previous investigations into the history of HCV epidemic have typically used a coalescent-based method called the skyline plot (Pybus et al., 2000; Strimmer & Pybus, 2001), which converts an observed virus genealogy into a plot of the effective number of infections through time. Importantly, the skyline plot does not require a model of demographic or epidemic history to be specified a priori, which is of benefit because using incorrect demographic models can lead to biased results. However, previously used skyline plot methods have a disadvantage – they infer demographic history from a single estimated genealogy, rather than from the sampled gene sequences, and thus ignore the error associated with phylogenetic reconstruction, which may be large. Here, we take advantage of a recently developed and more powerful method, called the Bayesian skyline plot (BSP) (Drummond et al., 2005), which correctly includes phylogenetic error. In essence, BSP infers demographic history directly from the sampled gene sequences, by ‘averaging’ across all possible genealogies, weighting each genealogy by its probability of being correct.

**METHODS**

**HCV subtype 1b sequences from China.** Partial E1 and NS5B HCV sequences from nine cities in China have been previously obtained (Fig. 1; Table 1) (Lu et al., 2005). Briefly, RNA-positive serum samples were collected in January 2002 by the laboratories of Da-An Diagnostic Center, Sun Yat-sen University, China. A total of 148 positive samples were randomly chosen. Partial E1 and NS5B gene regions were amplified by nested RT-PCR and the purified DNA was sequenced in both directions. In total, 89 partial E1 sequences and 92 partial NS5B sequences belonging to subtype 1b were obtained. A previous study by Lu et al. (2005) considered only 1b sequences from China and did not include 1b sequences from other countries. It showed that many of the Chinese 1b strains obtained from nine cities fell into two phylogenetic groups: 35 E1 sequences were in group A and 26 E1 sequences were in group B. Similarly, 35 NS5B sequences were in group A and 30 NS5B sequences were in group B (Fig. 1; Table 1) (Lu et al., 2005). The E1

**Table 1. Number and location of the HCV subtype 1b sequences analysed in this study**

When the numbers were different for the E1 and NS5B gene region, numbers for the NS5B region are shown in parentheses.

<table>
<thead>
<tr>
<th>City</th>
<th>Cluster A</th>
<th>Cluster B</th>
<th>Not in either cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenyang</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Beijing</td>
<td>7</td>
<td>1 (2)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Hohhot</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Shanghai</td>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Zhengzhou</td>
<td>2</td>
<td>4 (5)</td>
<td>2</td>
</tr>
<tr>
<td>Guangzhou</td>
<td>7</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Shenzhen</td>
<td>1</td>
<td>7 (8)</td>
<td>5</td>
</tr>
<tr>
<td>Foshan</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Kunming</td>
<td>0</td>
<td>0 (1)</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>26 (30)</td>
<td>28 (27)</td>
</tr>
</tbody>
</table>
region represents nucleotide positions 615–914 and the NS5B region represents nucleotide positions 7914–8288 relative to the sequence M62321.

**HCV subtype 1b global reference sequences.** HCV sequences representing the same genome region as the Chinese sequences were collated from the Hepatitis virus database (http://s2as02.genes.nig.ac.jp/). These sequences were then genotyped by phylogenetic analysis in order to identify subtype 1b strains. A total of 304 partial E1 and 323 partial NS5B sequences belonging to subtype 1b were identified. Next, information from the database and from original publications was used to exclude sequences if: (i) multiple sequences from the same individual were present, (ii) multiple sequences obtained from individuals with known epidemiological linkage were present, (iii) sequences were not directly isolated from infected individuals, and (iv) the nationality of the sampled individuals was unknown. Furthermore, when sequences from the same country clustered together, only one sequence was selected and left as a representative of that cluster. This left 72 E1 subtype 1b sequences from 25 countries and 61 NS5B subtype 1b sequences from 22 countries, excluding Chinese sequences. The sequence accession numbers and countries of isolation are shown in Figs 2 and 3.

**Phylogenetic analysis.** These reference strains were aligned with our Chinese sequences; initial multiple alignments were calculated using CLUSTAL_X 1.81 (Thompson et al., 1997), and were subsequently adjusted by hand. Phylogenies were calculated from the sequence alignments using the neighbour-joining method based on genetic distances calculated using the Kimura two-parameter substitution model, as implemented in MEGA version 3 (Kumar et al., 2004) or PAUP 4.0 (Swofford, D. L. Sinauer Associates, Sunderland, MA) under an HKY substitution model with a gamma-distribution model of among site rate heterogeneity. To assess the reliability of the phylogenies, bootstrap resampling and reestimation were carried out 1000 times for the neighbour-joining trees.

**Coalescent-based inference of HCV population dynamics.** The past population dynamics of strains from group A and B were estimated from the HCV gene sequences using the BSP as implemented in the program Beast 1.2 (http://evolve.zoo.ox.ac.uk/beast/) (Drummond et al., 2002). The BSP uses a Markov Chain Monte Carlo (MCMC) procedure to sample the distribution of generalized skyline plots, given a set of sampled gene sequences, and then combines these plots to generate a posterior distribution of effective population size through time (or the effective number of infections, in the case of viral epidemics). The BSP provides credibility intervals (confidence limits) for effective population size at every point in time, back to the most recent common ancestor (MRCA) of the sampled sequences.

Six datasets were analysed: (1) group A E1 sequences (n=35), (2) group A NS5B sequences (n=35), (3) group B E1 sequences (n=26), (4) group B NS5B sequences (n=30), (5) group A E1 + NS5B concatenated sequences (n=35), and (6) group B E1 + NS5B concatenated sequences (n=26). The latter two datasets were constructed by simply joining together the E1 and NS5B sequences that had been obtained from the same patient. As a result, four patients (ZZ7, BJ564, SZ63 and KM10) were not represented in dataset 6, because they were only sequenced in the NS5B region. To generate results on a timescale of years, we used previously published estimates of HCV substitution rates for subtype 1b: 5.0 × 10⁻⁴ substitutions per site per year for NS5B and 7.9 × 10⁻⁴ substitutions per site per year for E1 (Pybus et al., 2001). These evolutionary rates have been shown to be reliable in estimating the timing of epidemiological events (Pybus et al., 2003).

BSPs were calculated using an HKY-substitution model with a gamma-distribution model of among site rate heterogeneity. For the concatenated datasets, we assumed that E1 and NS5B had the same transition/transversion ratio, but had different substitution rates and different distributions of among site rate heterogeneity. Each MCMC was run for 10 000 000 states and sampled every 1000 states. MCMC convergence, effective sample size and burn-in were monitored using Tracer 1.2 (http://evolve.zoo.ox.ac.uk/beast/). Lastly, linear regression was used to estimate the viral exponential growth rate from the BSP results.

**RESULTS**

**Phylogenetic analysis**

Fig. 2(a) shows the neighbour-joining phylogeny estimated from the E1 gene sequences (89 Chinese strains plus 72 reference strains). The 35 strains of group A and the 26 strains of group B form two reciprocally monophyletic clades within the tree, and neither cluster included reference sequences from any other country (Fig. 2b and c). The bootstrap values for strains of groups A and B were 79 and 33 %, respectively. Of the remaining 28 Chinese subtype 1b sequences, some formed small clusters and the rest were randomly distributed among the reference sequences from other countries. A similar result was observed within the NS5B sequence phylogeny, shown in Fig. 3(a) (92 Chinese strains plus 61 reference strains). Again, the 35 strains of group A and the 30 strains of group B formed two clusters that did not contain any non-Chinese reference strains (Fig. 3b and c). The bootstrap values for clusters A and B were 13 and 72 %, respectively. Thus, the Chinese clusters of strains from groups A and B were found in both gene regions with strong bootstrap support (>70 %) in one gene but not in the other.

To further investigate this variability in bootstrap scores, we concatenated the Chinese E1 and NS5B sequences (i.e. datasets 5 and 6) and then added 118 subtype 1b reference strains (from the database) that had been concatenated in the same way. In this analysis, which combined the phylogenetic signal from both genes, the bootstrap values for groups A and B were both 100 % (see Supplementary material in JGV Online). Thus, the two groups are highly supported when the two genes are combined, and the above-mentioned low bootstrap scores are the result of less phylogenetic information in the datasets from individual genes.

Strains of group A were found in all cities, except Hohhot in Inner Mongolia and Kunming in the southwest (Table 1). Although strains of group B were found in many cities, they were predominantly from three cities in the Pearl River Delta and from Zhengzhou (Table 1). These results strongly suggest that they represent two independent chains of HCV transmission that have occurred within China. Furthermore, since strains of groups A and B are not geographically restricted, the epidemic history of the groups may reflect the dynamics of HCV infection throughout China. The remaining Chinese strains of HCV, which were distributed among subtype 1b reference sequences from other countries, probably reflect multiple sporadic introductions into China of various subtype 1b strains from other countries.
Fig. 2. Phylogeny of Chinese subtype 1b strains and reference strains estimated from E1 gene sequences (a). Strains of groups A and B form two monophyletic clusters. Arrows show the position of the two Chinese clusters. Groups A and B are detailed in (b) and (c), respectively. The bootstrap values for the clusters are shown. Strains beginning with SY, BJ, M, SH, ZZ, SZ, FS and KM represent strains sampled from Shenyang, Beijing, Hohhot, Shanghai, Zhengzhou, Shenzhen, Foshan and Kunming, respectively. Strains beginning with GZ and GB were sampled from Guangzhou. The tree was rooted with two subtype 1a strains (M62321 and D10749; not shown). The scale bars are in units of nucleotide substitutions per site.
Fig. 3. Phylogeny of Chinese subtype 1b strains and reference strains estimated from NS5B gene sequences (a). Arrows show the position of the two Chinese clusters. Groups A and B are detailed in (b) and (c), respectively. See Fig. 2 legend for further details.
Coalescent-based inference of HCV population dynamics

Six datasets were analysed using the BSP. Fig. 4(a) shows the population dynamics of group A reconstructed from E1 and NS5B gene sequences (i.e. datasets 1 and 2). The date of the MRCA of group A was estimated to be 1969 from the E1 sequences and 1964 from the NS5B sequences (Table 2). In addition to similar MRCA dates, both genes show similar estimates of epidemic history; the effective number of infections through time increases approximately exponentially from about 1970 to 1990, after which the growth rate slows considerably (Fig. 4a). The estimates of current effective population size obtained from the E1 and NS5B datasets were also very similar (Table 2).

Fig. 4(b) shows the equivalent BSPs obtained for group B (i.e. datasets 3 and 4). The date of the MRCA of group B was estimated to be 1978 for the E1 sequences and 1968 for the NS5B sequences (Table 2). The date of the MRCA of group B are less similar to each other than dates of group A. This is because the E1 alignment (dataset 3) contains fewer samples than the NS5B alignment (dataset 4); in particular, the divergent strain ZZ7 is not present in the E1 alignment, thus reducing the estimated age of that dataset. Again, both genes show a population dynamic history in which the virus initially increases at a rapid exponential rate, although the growth rate appears more variable through time for the NS5B dataset. The growth rate then slows considerably around 1985 in both datasets (Fig. 4b). Lastly, the E1 and NS5B genes gave similar values for the estimated effective number of infections at present (Table 2).

The best estimate of the population genetic history of groups A and B was obtained by concatenating E1 and NS5B sequences obtained from the same patient (i.e. datasets 5 and 6). The validity of this approach was demonstrated by the fact that the E1 and NS5B datasets give very similar results for each group (Fig. 4). The BSP results for the concatenated datasets are shown in Fig. 5(a). As expected, the concatenated datasets produce plots with smaller confidence

Table 2. Estimated evolutionary parameters for each dataset (median and 95% credibility intervals)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
<th>Date of MRCA (year)</th>
<th>Transition/transversion rate ratio ($\lambda$)</th>
<th>E1 gene rate heterogeneity parameter ($\alpha$)</th>
<th>NS5B gene rate heterogeneity parameter ($\alpha$)</th>
<th>Current effective no. infections</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Group A, NS5B gene</td>
<td>1964 (1949–1977)</td>
<td>12.5 (6-5–21-1)</td>
<td>NA</td>
<td>0-043 (0-000-0-097)</td>
<td>5070 (930–132 000)</td>
</tr>
<tr>
<td>4</td>
<td>Group B, NS5B gene</td>
<td>1968 (1953–1977)</td>
<td>12-3 (6-7–20-0)</td>
<td>NA</td>
<td>0-130 (0-001–0-220)</td>
<td>6760 (1070–193 000)</td>
</tr>
<tr>
<td>5</td>
<td>Group A, concatenated</td>
<td>1968 (1951–1978)</td>
<td>8-9 (6-2–12-0)</td>
<td>0-240 (0-140–0-380)</td>
<td>0-042 (0-000–0-099)</td>
<td>5800 (1140–163 000)</td>
</tr>
</tbody>
</table>

NA, Not applicable.
with the molecular epidemiological survey results (Lu et al., 2005), which showed that the proportions of sampled Chinese patients infected with strains of group A and B are approximately equal.

**DISCUSSION**

Our phylogenetic analysis indicates that there are two patterns of HCV subtype 1b transmission in China. The first pattern is represented by groups A and B, and reflects the ongoing and sustained transmission of particular viral lineages within the country, possibly through a shared transmission route. Other small clusters were found in the phylogenies, so further sampling of HCV from other regions in China may reveal other transmission chains of a similar nature. China has strictly restricted the importation of blood products (Shan et al., 2002), providing one possible explanation for the existence of the Chinese-specific transmission chains. The second pattern is represented by the individual Chinese isolates of HCV that are more closely related to non-Chinese isolates of HCV, which most likely represent multiple, sporadic migration of strains of HCV. We tried to investigate which countries these sporadic strains were derived from but our phylogenies were insufficiently resolved to answer this question reliably (data not shown). The introduction of such strains may arise through travel or immigration (Bernier et al., 1996; Morice et al., 2001; Stuyver et al., 1995), by the import and export of contaminated blood products (Kinosita et al., 1993) or through networks of injecting drug users (Cochrane et al., 2002). It is interesting to note that the two transmission patterns described above have also recently been noted for another virus; in an analysis of HIV-1 subtype B in the UK, Hue et al. (2005) found at least six co-circulating transmission clusters within a single risk group.

The observation that strains of groups A and B are geographically widespread and infect people throughout China requires explanation. In addition, the estimated exponential growth rates of the groups are unusually high. Previous skyline plot analyses have suggested lower rates of spread for subtype 1b (Nakano et al., 2004; Pybus et al., 2001; Tanaka et al., 2002); however, this is probably because previous estimates have represented the average worldwide growth of subtype 1b since its MRCA, whereas our analysis considers particular subtype 1b strains spreading more recently within a single population. Hue et al. (2005) also obtained fast rates of growth for individual HIV-1 transmission clusters. Furthermore, the growth rate of group A is similar to that estimated for HCV genotype 4 in Egypt (Pybus et al., 2003; Tanaka et al., 2004), which was spread rapidly during the twentieth century by extensive Egypt-wide anti-schistosomiasis injection campaigns (Frank et al., 2000). The growth rate of group B appears to be higher than that of HCV genotype 4 in Egypt.

According to Chen et al. (2002), the prevalence of anti-HCV in China is significantly higher among post-transfusional
hepatitis patients, haemodialysis patients and injecting drug users. Chen et al. (2002) also reported that anti-HCV prevalence is higher in plasma donors and paid blood donors than in whole blood donors and volunteer blood donors. It is therefore possible that the rapid spread and geographical dissemination of strains of groups A and B throughout China are related to iatrogenic transmission and/or past socio-medical conditions in China.

A cooperative medical system was developed in China in the 1950s to provide basic health care and preventive services (Hesketh & Wei, 1997). Widespread immunization campaigns were started at this time. However, medical schools and specialist hospital departments closed during the ‘Cultural Revolution’ during the years 1966–1976. Over a million non-professional health-care providers (‘barefoot doctors’) were provided with limited training and were responsible for the health care in the countryside (Brown et al., 1984; Hesketh & Wei, 1997). Under such circumstances there is potential for the transmission of blood-borne viruses such as HCV, particularly by unsafe injection and traditional Chinese acupuncture (Hsu, 1996). This period coincides with the origin of group A, and it is reasonable to hypothesise that the early spread of strains from group A was aided by transmission via parenteral routes.

Standard medical training was restarted in 1977 and trained physicians and nurses were reintroduced into the health-care system (Brown et al., 1984). Chinese hospitals were modernized considerably during the 1980s (Henderson et al., 1987), and blood transfusions would be expected to increase as hospitals modernize. However, a chronic shortage of blood led to a market for paid blood donations and associated illegal practices (Shan et al., 2002). For example, a 1985 outbreak of hepatitis C among plasma donors in Hebei province was linked to cross-contamination during plasma donation (Meng et al., 1991). Unlicensed centres for collection of plasma and whole blood increased in many parts of China until the early 1990s (Shan et al., 2002). Unsafe practices such as the reuse of non-sterile needles and the pooling and returning of blood from different donors have been reported in these centres (Shan et al., 2002). The dramatic growth rate of strains from group B could therefore result from the increased use of blood transfusion and blood products and an inadequate collection process after the Cultural Revolution, and the estimated origin of group B in 1975 agrees with this. It is likely that this transmission route also contributed to the spread of strains from group A during the 1980s.

Blood transfusion services and blood safety have improved in China since the discovery of HCV in 1989 (WHO, 2004). In 1998, the Chinese government banned paid whole-blood donations for clinical use and encouraged voluntary donation (WHO, 1999). Our analysis shows the effect of these improvements; the growth rates of both groups slowed considerably after 1990. HCV infection in injecting drug users has been well-studied in Yunnan, Sichuan, Guanxi, Xinjiang provinces in Southern China (Garten et al., 2004; Ruan et al., 2004; Zhang et al., 2002; Zhang et al., 2004), where infection has increased among young injecting heroin users since 1990. However, our analysis contained only one strain from this area, and the recent decrease in growth rates of groups A and B suggests that our results do not reflect recent HCV transmission among injecting drug users.

Hepatocellular carcinoma (HCC) is the fourth most common cause of death from cancer in China, and China alone accounts for 53% of all liver-cancer deaths worldwide (Pisani et al., 1999). In China, hepatitis B virus (HBV) infection is thought to be the main causative agent of HCC and the prevalence of HBV surface antigen in HCC patients is 63–84% (Deng et al., 1998; Yang et al., 2004; Zhang et al., 1998). Chronic HCV infection is also a risk factor for HCC in the decades following the initial infection (Seeff, 1997), and the prevalence of anti-HCV in HCC patients is 8–38% (Deng et al., 1998; Wang et al., 1999; Yang et al., 2004; Zhang et al., 1998). According to our analysis, the number of infections of groups A and B increased exponentially between 1970 and 1990. This suggests that, in the absence of treatment, the incidence of HCC resulting from infections of groups A and B will continue to increase dramatically during the next two decades. However, interferon therapy is reported to be less effective in patients infected with genotype 1 HCV than in patients infected with non-genotype 1 HCV (Zein et al., 1996). The future HCC disease burden could thus pose a significant threat to China’s public health and health-care system. Further epidemiological surveillance and continued efforts to treat and prevent HCV infection in China are therefore essential.

Our analysis used a recently developed method, the BSP, which has several advantages over previous methods. First, is that the BSP includes confidence limits for the estimated effective population size. Unlike previous skyline plots (implemented in the program GENIE; Pybus & Rambaut, 2002), the BSP takes into account all sources of statistical error and thus produces more accurate estimates and confidence limits. Second, is that the BSP can combine sequences from different genome regions, even if they have different rates of evolution, as was performed for the datasets 5 and 6. Third, is that the BSP provides a more accurate estimate of the age of the MRCA. A preliminary analysis of our data indicated that previous methods (i.e. GENIE) were sensitive to the presence of divergent phylogenetic branches at the root of the tree, the lengths of which were poorly supported by the data (results not shown). Last, is that the BSP allows us to detect novel demographic patterns that are not readily described by simple demographic models. Here, the BSP revealed a recent slowdown in transmission for groups A and B, which was not detected using the simple models implemented with GENIE. We hope that the BSP will continue to prove a useful tool in the molecular epidemiology of viruses.

**ACKNOWLEDGEMENTS**

O. G. P. was funded by the Royal Society.
REFERENCES


Morice, Y., Roulot, D., Grando, V. & 8 other authors (2001). Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. J Gen Virol 82, 1001–1012.


http://vir.sgmjournals.org


