Molecular Analysis of Hepatitis C Virus Infection in Bulgarian Injecting Drug Users

Massimo Ciccozzi,[1] Gianni Zehender,[2] Valeria Cento,[3] Alessandra Lo Presti,[1] Pavel Teoharov,[4] Ivan Pavlov,[4] Violeta Bogdanova,[5] Carlo Federico Perno,[3,6] and Marco Ciotti[6]*

[1]*Department of Infectious, Parasitic and Immunomediated Disease, Istituto Superiore di Sanita', Rome, Italy*
[2]*Department of Clinical Sciences, L. Sacco, University of Milan, Milan, Italy*
[3]*Department of Experimental Medicine and Biochemical Science, University of Rome Tor Vergata, Rome, Italy*
[4]*National Center of Infectious and Parasitic Diseases, Sofia, Bulgaria*
[5]*National Center for Addictions, Sofia, Bulgaria*
[6]*Laboratory of Molecular Virology, Foundation Polyclinic Tor Vergata, Rome, Italy*

Intravenous drug users constitute a group at risk for hepatitis C virus (HCV) infection. Today, no data are available on the molecular epidemiology of HCV in Bulgaria despite the fact that in recent years the incidence of acute hepatitis C infection among Bulgarian intravenous drug users increased sixfold and about 2/3 of them developed a chronic infection. The aim of this study was to determine the circulation of hepatitis C genotypes among drug users and to study the evolution and transmission history of the virus by molecular clock and Bayesian methods, respectively. Sequencing of NS5B gene showed that the genotype 3a was the most prevalent type among intravenous drug users. In the Bayesian tree, the 3a subtypes grouped in one main clade with one small cluster well statistically supported. The root of the tree was dated back to the year 1836, and the main clade from Bulgaria was dated 1960. The effective number of infections remained constant until about years 1950s, growing exponentially from the 1960s to the 1990s, reaching a plateau in the years 2000. The not significant intermixing with isolates from other countries may suggest a segregated circulation of the epidemic between 1940s and 1980s. The plateau reached by the epidemic in the early 2000s may indicate the partial success of the new preventive policies adopted in Bulgaria. *J. Med. Virol. 83:1565–1570, 2011.*
© 2011 Wiley-Liss, Inc.

KEY WORDS: hepatitis C virus; molecular epidemiolgy; evolution history; transmission history

INTRODUCTION

Hepatitis C virus (HCV) is the major cause of chronic hepatitis, cirrhosis of the liver and hepatocellular carcinoma in developed countries. About 170 million people are infected chronically with HCV worldwide. In recent years, the HCV epidemics worsened due to the virus dissemination by blood transfusions, blood-derived products, and unsafe medical practice. In industrialized countries, needle sharing is the major risk factor of viral dissemination among injecting drug users [Trepo and Pradat, 1999]. HCV is classified by phylogenetic methods into six major genotypes and many subtypes which differ in geographical distribution, and transmission route [Kuiken et al., 2009]. Genotype 1 and subtypes 1a and 1b are the most prevalent worldwide, though subtypes 3a and 1a are highly prevalent among injecting drug users [Kuiken et al., 2009]. Few reports are available on the molecular epidemiology of HCV infection among injecting drug users in Eastern Europe [Kalinina et al., 2001; Krekulova et al., 2001, 2005] and none on injecting drug users in Bulgaria. Therefore, a molecular analysis of HCV from drug users infected chronically was carried out to determine the epidemiology of the viral subtypes. Indeed, it

has been reported that some geographical differences may exist as reported in Czech Republic, where genotype 1b is the most prevalent subtype among drug users [Krekulova et al., 2001]. In this study, the transmission history of HCV in Bulgaria was investigated by Bayesian statistical inference framework to have the simultaneous reconstruction of the temporal and spatial history of the epidemic based on isolates randomly sampled at known times in different places [Lemey et al., 2009]. A comprehensive phylodynamic [Grenfell et al., 2004] study of HCV diversity in Bulgaria may also provide information about the demographic, social, and biological factors that have given rise to the current epidemiological patterns in this geographical area.

MATERIAL AND METHODS

Study Group

Serum samples were obtained from 32 drug users infected chronically with hepatitis C and admitted at the National Center for Addictions, Sofia, Bulgaria. After collection, the serum was stored at $-80°C$ until analysis.

HCV Genotyping

HCV RNA was extracted from 140 μl serum samples with the QIAamp Viral RNA Mini kit (Qiagen, Milan, Italy), then 10 μl of RNA were reverse transcribed by random examer and cDNA amplified by a heminested PCR with primers targeting the NS5B as reported in Laperche et al. [2005].

The amplified products were purified with the QIAquick PCR purification kit (Qiagen) and sequenced using the fluorescent dye terminator technology (BigDye Terminator v3.1 Cycle Sequencing kit; Applied Biosystem, Foster City, CA) according to the manufacturer's instructions and run on an ABI 3100 genetic analyzer (Applied Biosystems). Sequences were aligned using CLUSTAL X software [Thompson et al., 1994], then manually edited with the Bioedit software [Hall, 1999]. HCV genotype was assessed by phylogenetic analysis of the NS5B sequences. The ModelTest [Posada and Buckley, 2004] was used to select the simplest evolutionary model that fitted adequately the sequence data.

Hepatitis C Virus Dataset

Two hepatitis C data sets were built. The first one contained reference sequences downloaded from the Los Alamos HCV sequence database (http://hcv.lanl.gov/content/index) on the basis of the following inclusion criteria: (1) sequences already published in peer-reviewed journals, except the new sequences described below; (2) no uncertainty about the subtype assignment; (3) known city/state of origin as established in the original publication. Nine reference sequences related to the HCV genotypes and subtypes 1a, 1b, 3a, 3b, 4a, and 6a (AB444473, 157781216,

AB444480, AB444481, AB444482, AF169005, D90208, EU255952, FJ469164) were selected and added to the newly generated NS5B HCV gene sequences from Bulgaria. The second data set contained sequences from the most representative genotype of this study (genotype 3a) including sequences isolated from Western, Central, and East European countries. The sampling locations of the isolates were Bulgaria (n = 13); Cyprus (n = 8); Estonia (n = 33); Austria(n = 22); France (n = 58); Ireland (n = 1); Lithuania (n = 22); Spain (n = 4).

Evolutionary Rate Estimates and Time-Scaled Phylogeny Reconstruction

A statistical method, based on coalescent theory, was used to estimate the epidemic history of HCV subtype 3a from the viral gene sequences obtained as well as to describe the relationship between the demographic history of a population and the genealogy of individuals sampled randomly from it [Holmes et al., 1995; Pybus et al., 2000; Pybus and Rambaut, 2002].

As sampling dates were not available for all sequences included in the analysis, the evolutionary rates were estimated using a data set containing the same 342 nt NS5B fragment following the external calibration approach used previously by other authors [Pybus et al., 2009]. The phylogeny obtained from the original data set was calibrated by adjusting the mean substitution rate to the mean external evolutionary rate estimates. The external sequences sample included 148 dated HCV NS5B sequences of genotype 3a from Western, Central, and Eastern European countries, which were retrieved from the Los Alamos HCV Sequence Database, as well as the new Bulgarian isolates. The sampling dates ranged from 2000 to 2009.

Evolutionary rates were estimated using a Bayesian–Markov–Chain–Monte–Carlo (MCMC) method implemented in BEAST package 1.5.3 [Drummond et al., 2005; Drummond and Rambaut, 2007] and both a strict and relaxed clock with an uncorrelated log normal rate distribution. As coalescent priors, three parametric demographic models of population growth (constant size, exponential, and logistic growth) and a Bayesian skyline plot (BSP, a non-parametric piecewise-constant model) were compared. Analysis was performed as follows. To reconstruct the time-scaled phylogeny of the main data set, the same Bayesian MCMC method [Drummond et al., 2005; Drummond and Rambaut, 2007], molecular clock and demographic models, assuming the GTR + G + I model of nucleotide substitution, were used. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade.

The MCMC chains were run for at least 50 million generations, and sampled every 5,000 steps. Convergence was assessed on the basis of the effective sampling size (ESS) after a 10% burn-in [Drummond and Rambaut, 2007] using Tracer version 1.5 (http://

tree.bio.ed.ac.uk/software/tracer/). Only ESS values of >250 were accepted. Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals, and the best fitting models were selected using a Bayes factor (BF, using marginal likelihoods) implemented in BEAST [Drummond and Rambaut, 2007]. According to Suchard et al. [2001], the strength of the evidence against H_0 was evaluated as follows: 2lnBF <2 = no evidence; 2–6 = weak evidence; 6–10 = strong evidence; and >10 = very strong evidence. A negative 2lnBF indicates evidence in favor of H_0. Only values of ≥ 6 were considered significant.

Population dynamics were also analyzed in the Bulgarian clade on an individual basis by comparing the three coalescent models (constant, exponential, and BSP) and implementing a relaxed molecular clock model under the conditions described above. This analysis was not applied to genotype 1a because of the small sample size (six sequences).

RESULTS

Of the 32 injecting drug users examined, NS5B gene sequences were obtained from 20 of them. The maximum-likelihood tree showed that 13 patients were infected by 3a subtype, one by 1b subtype and six by 1a subtype, Figure 1. The distribution of the different subtypes by age, sex, mode of transmission, residence is reported in Table I.

In Figure 2 are shown the genotype 3a sequences analyzed to infer the demographic history in Bulgaria. The reconstruction of a Bayesian phylogenetic tree of HCV 3a subtype, in a calendar timescale, showed that the majority of the Bulgarian strains (11 sequences) segregated together and, with the exception of only three sequences from France, formed a highly significant clade (posterior probability > than 90%). Two sequences clustered instead with sequences from Cyprus. The molecular clock analysis estimated that the HCV 3a subtype was present in Bulgaria since the early 1960s for the main clade, with a lower limit estimate for the first introduction around the 1940s. Two sequences only suggested a more recent introduction of the virus around the end of the 60s. Based on the approximate marginal likelihoods of the six different demographic models used, the BF favored models enforcing a relaxed molecular clock over strict clock models. Models assuming exponential population growth performed always better than models assuming constant population size. In contrast, the BF was not significant when exponential models were compared with BSPs (data not shown) suggesting that both the parametric and the non-parametric models fit the data equally well. The tMRCA of all internal nodes was estimated using a relaxed molecular clock model. The root of the tree was dated back to the year 1836 (95%HPD 1743–1933), the main clade from Bulgaria was dated to 1960, (95%HPD 1948–1980), whereas the small cluster was dated to 1968 (95%HPD 1955–1987), Figure 2. The comparison
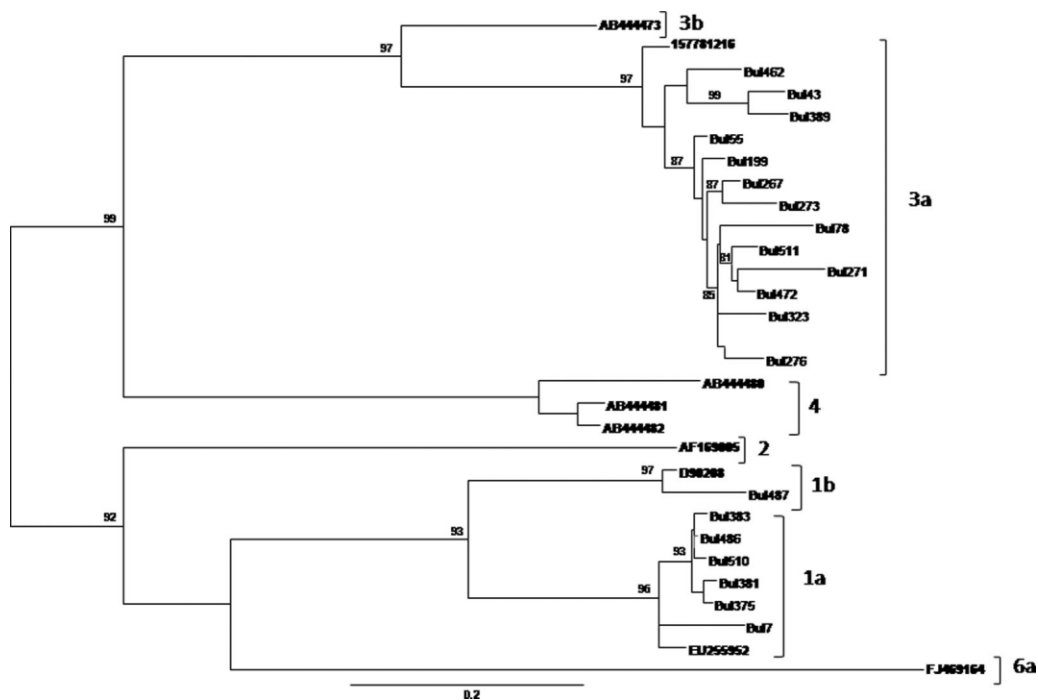

Fig. 1. HCV genotype was assessed by phylogenetic analysis of NS5B sequences. The phylogenetic analysis was performed on the 20 sequences of HCV-infected patients included in the study, adding to the alignment 9 reference strains for HCV genotypes 1a, 1b, 2, 3a, 3b, 4, 4a, and 6a sub-genotypes (AB444473, 157781216, AB444480, AB444481, AB444482, AF169005, D90208, EU255952, FJ469164).

TABLE I. Distribution of the Different HCV Subtypes by Age, Sex, and Mode of Transmission

| Patient	HCV genotype	Sex	Age	Transmission	City
462	3a	Male	28	Intravenous drug use	Botevgrad
43	3a	Male	28	Intravenous drug use	Sofia
389	3a	Female	19	Intravenous drug use	Sofia
55	3a	Female	23	Intravenous drug use	Burgas
199	3a	Female	30	Intravenous drug use	Sofia
267	3a	Male	26	Intravenous drug use	Sofia
273	3a	Female	25	Intravenous drug use	Sofia
78	3a	Male	25	Intravenous drug use	Radnevo
511	3a	Male	42	Intravenous drug use	Sofia
271	3a	Male	22	Intravenous drug use	Pasargik
472	3a	Female	24	Intravenous drug use	Sofia
323	3a	Male	23	Intravenous drug use	Pleven
276	3a	Male	29	Intravenous drug use	Sofia
487	1b	Male	57	Intravenous drug use	Sofia
383	1a	Male	26	Intravenous drug use	Sofia
486	1a	Female	39	Intravenous drug use	Sofia
510	1a	Male	25	Intravenous drug use	Sofia
381	1a	Male	34	Intravenous drug use	Sofia
375	1a	Male	29	Intravenous drug use	Plovdiv
7	1a	Male	46	Intravenous drug use	Svoge
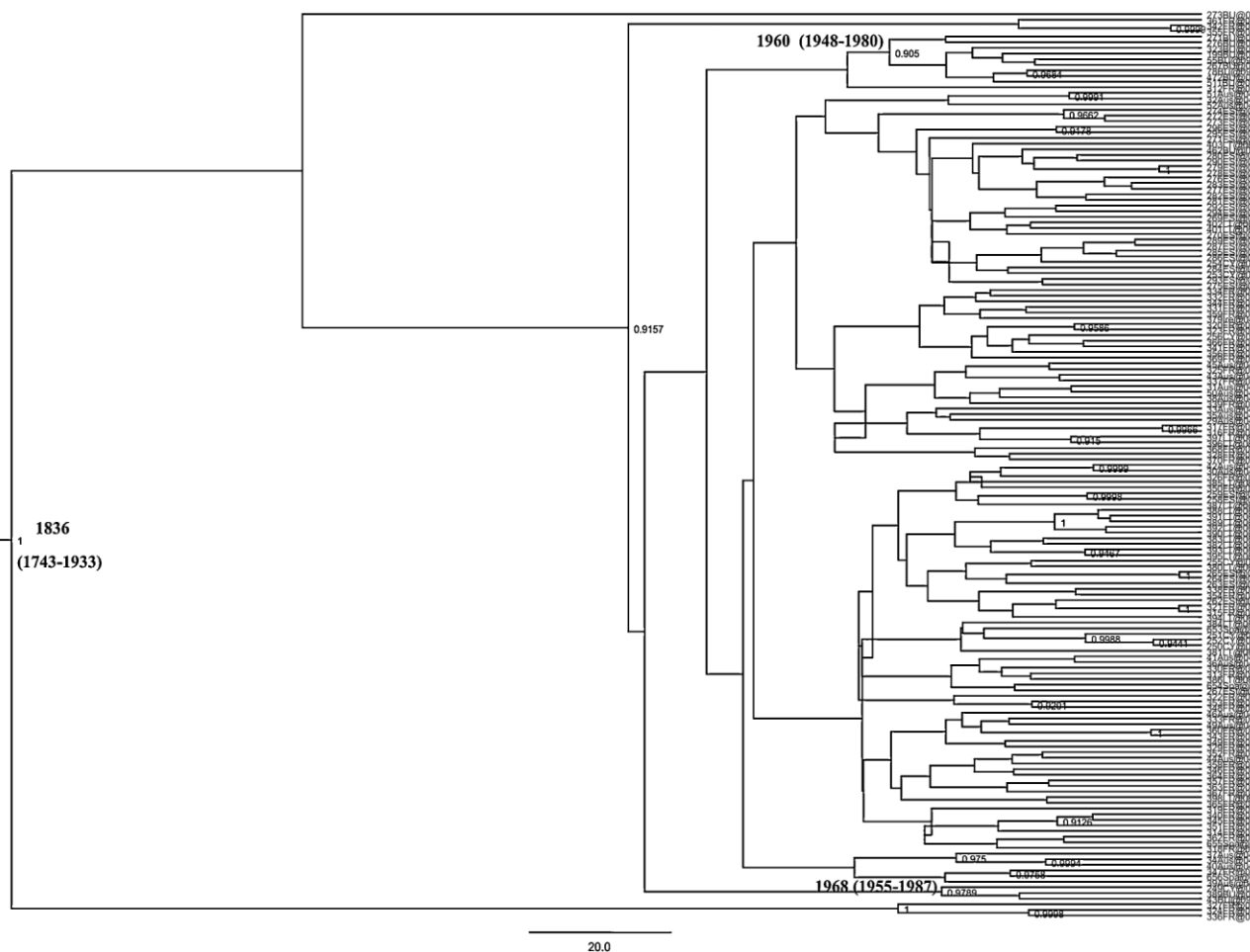
Fig. 2. Bayesian time-scaled tree of Bulgarian and European HCV NS5B genotype 3a sequences. The tMRCA, with the credibility interval based on 95% highest posterior density interval (HPD), of the root and of the two statistically supported (posterior probability > 1) internal nodes was reported in years.

between parametric and non-parametric models by BF showed that the BSP was preferred. The BSP showed that the effective number of infections remained constant until about the 1950s, growing exponentially from the 1960s to the 1990s, and reaching a plateau in the years 2000, Figure 3.

DISCUSSION

Hepatitis C chronic infection is a global health problem which affects about 3% of the world population [http://www.who.int/csr/disease/hepatitis/whocdscsrlyo2003/en/index1.html]. HCV genotypes 1, 2, and 3 constitute more than 90% of world chronic infections. Among them, subtypes 1a and 3a are common in injecting drug users [Kuiken et al., 2009], although specific geographical situations have been reported [Krekulova et al., 2005]. A detailed investigation of the molecular epidemiology of HCV among Bulgarian injecting drug users is described.

The molecular monitoring showed that the circulation of HCV subtypes in this risk group reflects that reported by a number of groups worldwide [Kuiken et al., 2009]. Analysis of the NS5B gene of the HCV 3a subtype, the most prevalent subtype identified in this study, allowed us to reconstruct the epidemiological history of HCV 3a subtype in Bulgaria. The Bulgarian sequences were analyzed together with those obtained from neighboring Mediterranean or Eastern countries. The phylogenetic, phylodinamic, and coalescent-based analyses provided a clear picture of its history. Indeed, Bayesian methods have been used widely because of their ability to integrate information about mean evolutionary rates, dated phylogeny and coalescent population dynamics.

The evolutionary rate of the NS5B gene detected in Bulgarian injecting drug users was estimated using an external data set aligned and cropped with the Bulgarian sequences, as described previously [Pybus et al., 2009]. It was found that it fell between previously calculated values [Power et al., 1995; Salemi and Vandamme, 2002; Tanaka et al., 2002; Magiorkinis et al., 2005, 2009; Ciccozzi et al., 2011].

It was estimated that HCV 3a subtype was present in Bulgaria since the early 1960s even though two sequences suggested the possibility of a more recent introduction of the virus around the end of the 60s.

The coalescent-based population dynamics analysis of the HCV 3a subtype done using a non-parametric coalescent model (BSP), made possible to estimate past changes in effective population size [Drummond et al., 2005]. The analysis showed a 2-log increase in the effective number of HCV 3a subtype infections between the end of 1940s (when the exponential growth began), and the end of 1990s, when the plot reached a plateau that still persists today. However, this result does not reflect probably the specific course of the Bulgarian outbreak, but also that of the HCV subtype
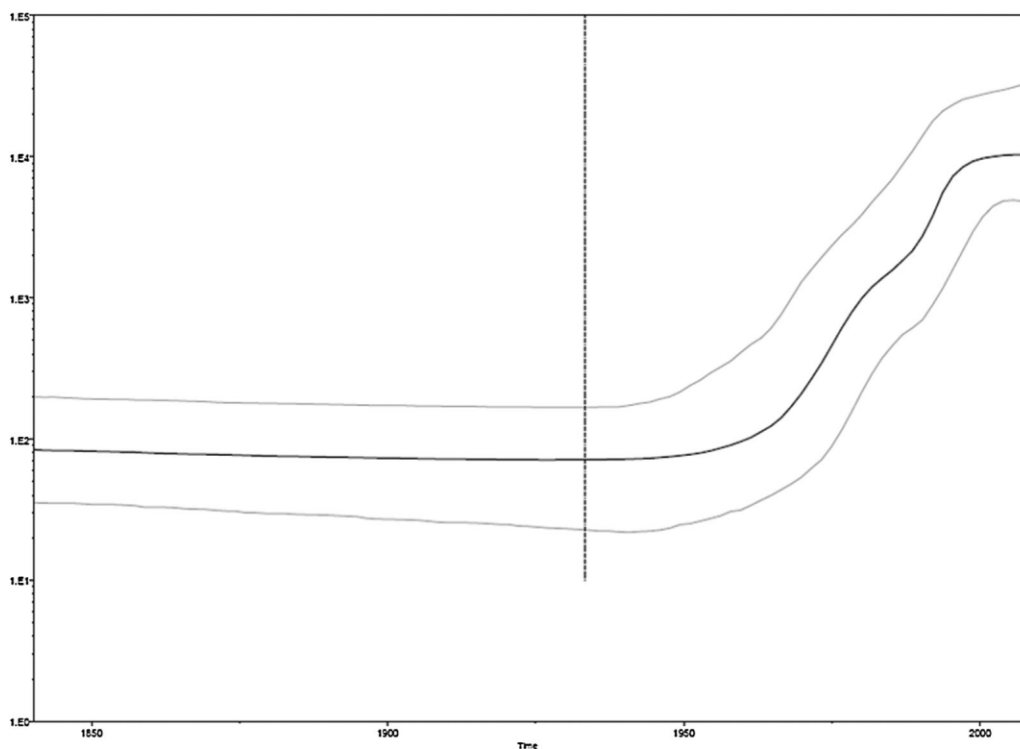

Fig. 3. Bayesian skyline plot (BSP) of Bulgarian and European HCV NS5B genotype 3a sequences. The effective number of the infections is reported on the Y axis. Time is reported in the X axis. The grey area corresponds to the credibility interval based on 95% highest posterior density interval (HPD).

3a in Europe, as reported by Pybus et al. [2005]. Furthermore, this period coincides with a considerable increase of the incidence of HCV infection especially among injecting drug users (46%) in Bulgaria as estimated in 2000–2006 [Boykinova et al., 2009]. Thus, the estimate seems to reflect the growth of the injecting drug users risk group rather than the innate transmission potential of the virus. Therefore, it cannot be excluded that the study population is not truly representative of all HCV cases in Bulgaria. Based on this analysis, the growth of the epidemic was relatively explosive and associated with the expansion of intravenous drug use [Magiorkinis et al., 2009; Pybus et al., 2005].

In conclusion, the Bayesian framework, which allows a spatial-temporal reconstruction of a phylogeny and an estimate of the population dynamics using a coalescent-based approach, indicates that the HCV 3a subtype epidemic entered in Bulgaria first in the 1960s as a result of a main ancestral event involving most of the cases. The slight and not significant intermixing with isolates from other countries may suggest a segregated circulation of the epidemic in the country between 1940s and 1980s, due probably to unsafe parenteral medical procedures but with drug addiction playing a relatively important role. The plateau reached by the epidemic in the early 2000s seems to indicate the partial success of the new preventive policies adopted in Europe and Bulgaria.

REFERENCES

Boykinova OB, Stoilova YD, Tsvetkova TZ, Baltadjiev IG. 2009. Epidemiological, immunological and clinical characteristics of acute hepatitis C. Folia Med 51:61–69.

Ciccozzi M, Ciccaglione AR, Lo Presti A, Yalcinkaya T, Pinar Taskan ZP, Equestre M, Costantino A, Bruni R, Ebranati E, Salemi M, Gray R, Rezza G, Galli M, Zehender G. 2011. Reconstruction of the evolutionary dynamics of the Hepatitis C virus 1b epidemic in Turkey. Infect Genet Evol 11:863–868.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185–1192.

Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332.

Hall TA. 1999. Bioedit: A user-friendly biological sequence alignment, editor and analysis program for Windows 95/98 NT. Nucl Acids Symp Ser 41:95–98.

Holmes EC, Nee S, Rambaut A, Garnett GP, Harvey PH. 1995. Revealing the history of infectious disease epidemics through phylogenetic trees. Philos Trans R Soc Lond B Biol Sci 349:33–40.

Kalinina O, Norder H, Vetrov T, Zhdanov K, Barzunova M, Plotnikova V, Mukomolov S, Magnius LO. 2001. Shift in predominating subtype of HCV from 1b to 3a in St. Petersburg mediated by increase in injecting drug use. J Med Virol 65:517–524.

Krekulova L, Rehak V, Madrigal N, Johnson M, Killoran P, Riley LW. 2001. Genotypic and epidemiologic characteristics of hepatitis C virus infections among recent injection drug user and nonuser populations. Clin Infect Dis 33:1435–1438.

Krekulova L, Rehak V, Riley LW. 2005. Hepatitis C virus (HCV) 5_NC sequence variants and their association with hepatitis C risk groups. J Clin Virol 32:300–304.

Kuiken C, Simmonds P. 2009. Nomenclature numbering of the hepatitis C virus. Methods Mol Biol 510:33–53.

Laperche S, Lunel F, Izopet J, Alain S, Deny P, Duverlie G, Gaudy C, Pawlotsky JM, Plantier JC, Pozzetto B, Thibault V, Tosetti F, Lefrere JJ. 2005. Comparison of hepatitis c virus NS5b and 5' noncoding gene sequencing methods in a multicenter study. J Clin Microbiol 43:733–739.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. Plos Comput Biol 5:1–16.

Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, Pybus OG, Allain JP, Oliver G, Pybus Cochraneb A, Holmes EC, Simmonds P. 2005. The hepatitis C virus epidemic among injecting drug users. Infect Genet Evol 5:131–139.

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

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808.

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

Pybus OG, Rambaut A. 2002. GENIE: Estimating demographic history from molecular phylogenies. Bioinformatics 18:1404–1405.

Pybus OG, Rambaut A, Harvey PH. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics 155:1429–1437.

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

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. 2009. Genetic history of Hepatitis C virus in East Asia. J Virol 83:1071–1082.

Salemi M, Vandamme AM. 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. J Mol Evol 54:62–70.

Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18:1001–1013.

Tanaka Y, Hanada K, Mizokami M, Yeo AE, Shih JW, Gojobori T, Alter HJ. 2002. Inaugural Article: 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 99:15584–15589.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1994. CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882.

Trepo C, Pradat P. 1999. Hepatitis C virus infection in Western Europe. J Hepatol 31:80–83.