

Phylogenetic and Phylogeographic Patterns of the HIV Type 1 Subtype F1 Parenteral Epidemic in Romania

Jean L. Mbisa,¹ Stéphane Hué,² Andrew J. Buckton,¹ Richard E. Myers,¹ Dan Duiculescu,³ Luminita Ene,³ Cristiana Oprea,³ Gratiela Tardei,⁴ Sorin Rugina,⁵ Mariana Mardarescu,⁶ Corinne Floch,⁷ Gundula Notheis,⁸ Bettina Zöhrer,⁹ Patricia A. Cane,¹ and Deenan Pillay²

Abstract

In the late 1980s an HIV-1 epidemic emerged in Romania that was dominated by subtype F1. The main route of infection is believed to be parenteral transmission in children. We sequenced partial *pol* coding regions of 70 subtype F1 samples from children and adolescents from the PENTA-EPPICC network of which 67 were from Romania. Phylogenetic reconstruction using the sequences and other publically available global subtype F sequences showed that 79% of Romanian F1 sequences formed a statistically robust monophyletic cluster. The monophyletic cluster was epidemiologically linked to parenteral transmission in children. Coalescent-based analysis dated the origins of the parenteral epidemic to 1983 [1981–1987; 95% HPD]. The analysis also shows that the epidemic's effective population size has remained fairly constant since the early 1990s suggesting limited onward spread of the virus within the population. Furthermore, phylogeographic analysis suggests that the root location of the parenteral epidemic was Bucharest.

THE USE OF UNSTERILIZED NEEDLES and syringes in a healthcare setting led to the emergence of an HIV-1 epidemic in Romania involving approximately 10,000 babies and children in the late 1980s.^{1–4} A distinctive feature of the Romanian epidemic was that it was dominated by subtype F HIV-1, which accounts for less than 1% of the infections worldwide, with the majority being found in Africa and South America.^{5–8} Subtype F is further divided into two sub-subtypes, namely F1 and F2.⁹ The Romanian epidemic was caused by subtype F1. In Europe, subtype F is rarely observed outside of Romania suggesting that the Romanian F1 epidemic might have originated from elsewhere.^{10–12}

A recent study showed that F1 sequences sampled in Romania are closely related to those from Angola and dated the time of the most recent common ancestor (tMRCAs) of the Romanian sequences to 1978.¹² The study also revealed that the Romanian epidemic is composed of two distinct strains of viruses: one consisting of sequences sampled mainly from

children and the second consisting mostly of adult sequences. Here, we specifically examine in detail the phylogenetic and phylogeographic patterns of the Romanian parenteral epidemic in children using new sequences with linked epidemiological data collected over a 6-year period as part of a EuroCoord-CHAIN study. EuroCoord-CHAIN is a large European collaboration of HIV observational cohorts (EuroCoord), including the PENTA-EPPICC network (Paediatric European Network for Treatment of AIDS-European Pregnancy and Paediatric HIV Cohort Collaboration and the European Collaborative HIV and anti-HIV Drug Resistance Network (CHAIN)).

As part of the study we performed population-based sequencing of part of the *pol* gene of 176 samples from HIV-1-infected children and adolescents in the PENTA-EPPICC network, collected between 2002 and 2008. Seventy of the samples were from children infected with HIV-1 subtype F1, of which 67 were from Romania. To investigate the genetic

¹Virus Reference Department, Microbiology Services, Health Protection Agency, London, United Kingdom.

²MRC Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London, London, United Kingdom.

³AIDS Department, Victor Babes Hospital of Infectious and Tropical Diseases, Bucharest, Romania.

⁴Immunology, Virology, and Molecular Biology Department, Victor Babes Hospital of Infectious and Tropical Diseases, Bucharest, Romania.

⁵Hospital for Infectious Diseases, Constanta, Romania.

⁶Matei Bals National Institute of Infectious Diseases, Bucharest, Romania.

⁷AP-HP, Hôpital Louis Mourier, Service de Pédiatrie, Colombes, France.

⁸Childrens Hospital of the Ludwig Maximilians University, Munich, Germany.

⁹Medical University of Graz, Graz, Austria.

TABLE 1. GEOGRAPHIC DISTRIBUTION OF SUBTYPE F1 AND F2 SEQUENCES USED FOR THE PHYLOGENETIC RECONSTRUCTION

Country of sampling	Subtype F1		Subtype F2
	EuroCoord ^a	Los Alamos ^b	Los Alamos ^b
Romania	54	338	
Brazil		157	
Angola		22	
Belgium		7	2
Spain		4	2
France	1	4	1
Germany	1		1
Italy		24	
Portugal		3	
Czech Republic		5	
Russia		3	
Great Britain		2	
Luxembourg		1	
Finland		1	
Switzerland		1	
Austria	1		
Slovenia		1	
Argentina		3	
United States		3	
Cuba		1	
Honduras		1	
Cameroon		1	23
DRC		1	1
Equatorial Guinea			1
Republic of Congo		1	
Mozambique		1	
Total	57	585	31

^aSequences from the EuroCoord/PENPACT1 study generated by this study.

^bSequences downloaded from the Los Alamos HIV sequence database.

relationship of the subtype F1 sequences from Romania to worldwide subtype F sequences, we compiled a dataset that included 57 subtype F1 sequences we genotyped plus 616 subtype F1 and F2 sequences from different geographic regions of the world that were downloaded from the Los Alamos HIV sequence database (Table 1). Where multiple sequences from the same patient or identical sequences were present, only the most recent sequence was used for the analysis.

The sequences were aligned using the Clustal W algorithm in MEGA4¹³ and were manually edited and trimmed to encompass only the first 858 nucleotides of the reverse transcriptase region of the HIV-1 *pol* gene. A preliminary phylogenetic analysis of the sequence alignment was performed using a rapid parallelized maximum-likelihood (ML) inference method (RAxML)¹⁴ utilizing a general-time reversible (GTR) model of nucleotide substitution. This analysis showed that the sequences generally segregated into three main groups of subtype F1 sequences and one group of subtype F2 sequences (Fig. 1).

The first group of F1 sequences contained a majority of sequences from Romania (390 out of 411; in red in Fig. 1). The other two F1 groups contained a majority of sequences from Angola (19 out of 21; in green in Fig. 1) and Brazil (157 out of

197; in yellow in Fig. 1), respectively. The three F1 clusters had high bootstrap support values of 84%, 100%, and 75%, respectively, as did the F2 cluster that had a bootstrap value of 100%. A sequence from the Republic of Congo was at the base of the F1 and F2 clusters indicating that the origin of the global subtype F epidemic was in Central Africa, which concurs with previous observations.^{11,12} Thus, this initial analysis suggests that most of the Romanian F1 epidemic is composed of lineages that cluster separately from other worldwide subtype F1 lineages.

To investigate the evolution and origins of the Romanian subtype F1 epidemic further we performed a more robust phylogenetic reconstruction using Bayesian Markov chain Monte Carlo (MCMC) methods as implemented in the BEAST program version 1.6.1.^{15,16} To undertake this analysis, we reduced the number of sequences in our dataset to 128 using the following criteria: (1) sequences without drug resistance mutations or from treatment-naïve individuals, and (2) sequences with known year of sampling. The new dataset included 52 sequences from Romania, of which 38 were from the EuroCoord-CHAIN study.

An alignment of the dataset was then used for phylogenetic analyses using Bayesian MCMC inference under the Hasegawa, Kishino, and Yano (HKY) model of substitution. All the Bayesian MCMC analyses carried out in this study were preceded by a molecular clock and demographic model comparison using Bayes factor (with a BF >20 level of significance) to determine the appropriate models to use (data not shown). A Bayesian-inferred phylogeny was estimated from one MCMC run of 50,000,000 generations with sampling every 5,000th generation and a 10% burn-in under a relaxed molecular clock and constant population size. The resulting maximum clade credibility (MCC) tree showed three main clusters of F1 sequences each containing a majority of sequences from Romania, Angola, and Brazil (Fig. 2; Clusters I, II, and III, respectively) and a cluster of F2 sequences (Cluster IV) with posterior probabilities >0.9.

This clustering of F sequences was similar to that observed using the rapid parallelized ML method (Fig. 1). Of particular note, 79% (41 out of 52) of the Romanian F1 sequences were part of the highly supported monophyletic cluster (Cluster I) with a posterior probability of 1. The cluster included three non-Romanian sequences from Austria, Germany, and United States. Interestingly, all the sequences within the monophyletic cluster were from children, whereas the sequence at the base of the cluster was from an adult patient (labeled "a†" in Fig. 2). Furthermore, the ancestral node of the monophyletic cluster and the adult sequence at the base of the cluster were highly supported with a posterior probability of 0.99. The adult sequence was part of a group of nine sequences, eight of which were from Romania and one from France. Of the eight Romanian sequences, five were from adults.

This phylogenetic analysis suggests that a major bottleneck event might have occurred resulting in the establishment of the Romanian monophyletic cluster from one or a few closely related lineages. Data on the route of transmission were available for the samples within the EuroCoord-CHAIN study as well as some of the samples downloaded from the Los Alamos HIV sequence database.

When this information was mapped on the monophyletic cluster it showed that the route of transmission was parenteral for 37 out of 41 of the Romanian sequences in Cluster I. The

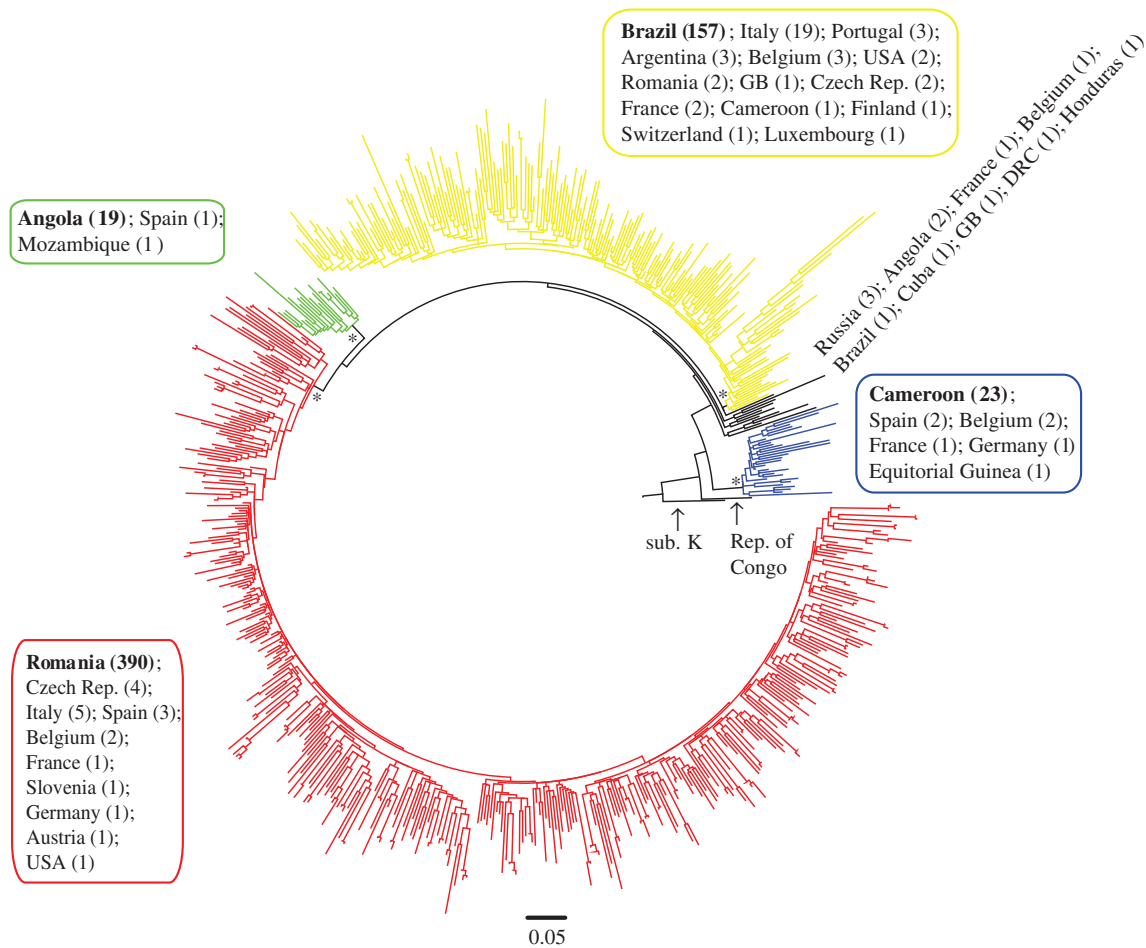


FIG. 1. Polar maximum-likelihood (ML) phylogenetic tree of global subtype F sequences. The phylogenetic tree was constructed by a rapid parallelized ML inference method using the general time reversible (GTR) model of substitution. The branches are drawn in scale with the bar at the bottom representing 0.05 nucleotide substitutions per site. The tree is rooted using a subtype K sequence (sub. K; accession no. AJ249239K). The high statistically supported nodes subtending the different geographic clusters are indicated by an asterisk (*), and outlined in different colors: red, majority of Romanian F1 sequences; green, majority of Angolan F1 sequences; yellow, majority of Brazilian F1 sequences; blue, subtype F2 sequences. The country of origin and the number of sequences represented in each cluster are shown next to the cluster outlined by a colored box matching that of the cluster.

routes of transmission for the remaining Romanian sequences in Cluster I were blood transfusion (1/41), perinatal (1/41), heterosexual contact (1/41), and unknown (1/41). In contrast, only two out of the remaining 13 Romanian sequences were transmitted parenterally and both sequences came from children. The remaining sequences from children in this group were acquired perinatally (2/5) or through an unknown route (1/5). In contrast, two of the adult sequences in the group were acquired sexually, whereas the remaining sequences had unknown routes of transmission. Taken together these data indicate that sequences in Cluster I represent the Romanian F1 parenteral epidemic in children and that the founder of Cluster I was probably a lineage introduced into the adult population in Romania and/or Europe at an earlier time.

The Bayesian-inferred phylogeny also showed that the Angolan sequences (Cluster II) were nested between a group of sequences from Romania and Europe suggesting that the Romanian sequences are more closely related to the Angolan than to the Brazilian sequences (Fig. 2.). This suggests an African origin for the HIV-1 subtype F1 lineage in Romania;

however, this observation had weak statistical support (posterior probability of 0.48 for the closest ancestral node linking the Angolan and Romanian lineages).

Next, the history of the parenteral F1 epidemic in Romania was reconstructed using the program BEAST. The tMRCA of the parenteral cluster was estimated to be 1983 [1981–1987; 95% highest probability density (HPD)]. Demographic reconstruction under a strict molecular clock and a Bayesian Skyline Plot (BSP) demographic model showed that the number of effective infections (N_e) for the parenteral epidemic, a number that represents infections contributing to new transmissions, grew exponentially until the early 1990s with an initial median estimate of 7 [2–38; 95% HPD] in 1983. However, the effective population size has remained constant from the early 1990s to 2008, the final year of sampling, with a final median estimate of 465 [193–1182; 95% HPD], which is approximately 5% of the estimated 10,000 children infected parenterally (Fig. 3). This suggests that very little onward spread through the parenteral route or by other means has occurred since the early 1990s. The analysis also revealed that

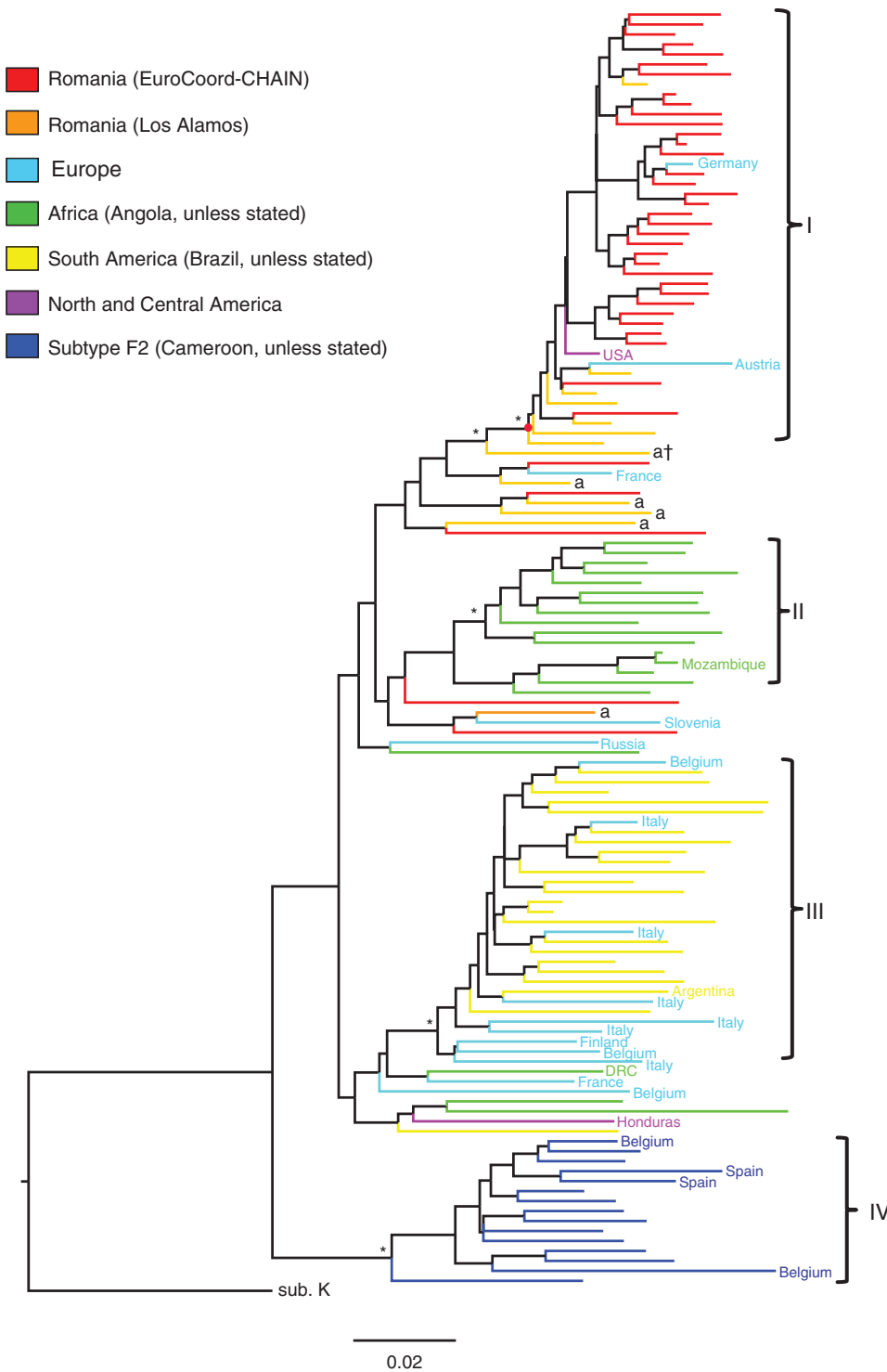


FIG. 2. Bayesian maximum clade credibility (MCC) tree of subtype F sequences. The Bayesian-inferred phylogeny of 128 global subtype F sequences. Branch lengths are drawn to scale with the bar at the bottom representing 0.02 nucleotide substitutions per site. The color of each tip branch represents the country or continent of origin of the sequence as shown in the figure legend. The country of origin of some of the sequences is also shown at the branch tips. Major ancestral nodes with significant statistical support (posterior probability >0.9) are indicated with an asterisk (*). Four significantly supported clusters are indicated by roman numerals: I, Romania monophyletic; II, Angola; III, Brazil; IV, F2 sequences. The red dot represents the time of the most recent common ancestor (tMRCA) of the Romania monophyletic cluster. Adult Romanian sequences are indicated with an “a” at the tip of the branches; a† indicates adult sequence at the base of the parenteral cluster. The tree is rooted using a subtype K sequence (sub. K; accession no. AJ249239K).

the parenteral epidemic viruses have evolved at a slightly slower rate compared to other HIV-1 epidemics and was estimated to be 1.1×10^{-3} [$0.7-1.5 \times 10^{-3}$; 95% HPD].¹⁷

Lastly, we estimated the phylogeographic pattern of the subtype F1 epidemic in Romania. Epidemiological data were available for the EuroCoord-CHAIN study as well as some of the sequences downloaded from the Los Alamos HIV sequence database, which included the county of origin. The data showed that the children were mostly infected in their

first year of life; therefore the county of birth of the child was ascribed as the origin of the sequences.

The sequences originated from 16 counties in Romania (Fig. 4). This information was used to estimate the geographic origin of the ancestral node of the parenteral epidemic using Bayesian inference framework, which models spatial diffusion of time-stamped lineages as a continuous-time Markov chain (CTMC) process over discrete sampling locations as implemented in BEAST software.¹⁸ Using a relaxed molecular

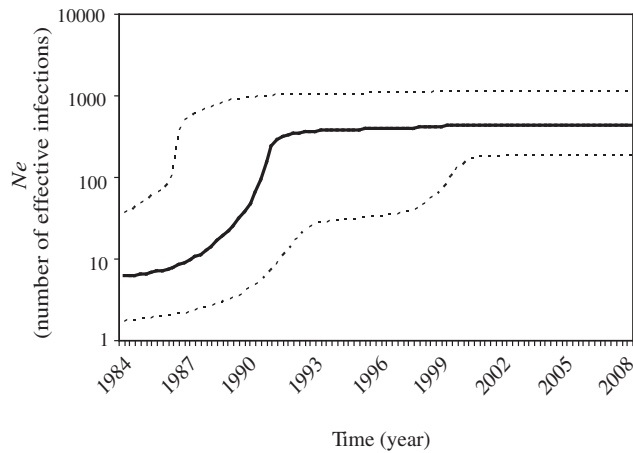


FIG. 3. Demographic history of the Romanian parenteral F1 epidemic. A Bayesian skyline plot of the Romanian monophyletic cluster (Cluster I). The black line represents the median and the dotted lines indicate the 95% lower and upper high posterior density (HPD) estimates of the effective population size (N_e).

clock model and a BSP demographic model we show that Bucharest is the location of the ancestral node subtending the parenteral cluster (root state probability of 0.47; Fig. 4). Only two other locations showed a root state probability greater than the prior one of 0.0625, these being Giurgiu (0.25) and Mures (0.08).

Of note, the rest of the parenteral cluster showed no clear spatial structure with the major ancestral nodes having undetermined root state locations. This finding was not surprising considering the route of transmission and the constant population dynamics. In contrast, the ancestral nodes at the base of the parenteral cluster that includes sequences sampled from the adult population indicate Bucharest as the predominant location. However, it is worth noting that these results could have been influenced by the high number of sequences from Bucharest in this group.

In summary, our findings make three important observations. First, we show that a large part of the Romanian subtype F1 epidemic is represented by a subpopulation of viruses sampled from children who mostly acquired the infection via the parenteral route as earlier as 1983, 5 years after the estimated introduction of subtype F1 into Romania.¹² This is also in keeping with the first reports of the HIV-1 F1 epidemic in Romanian children in 1989.¹

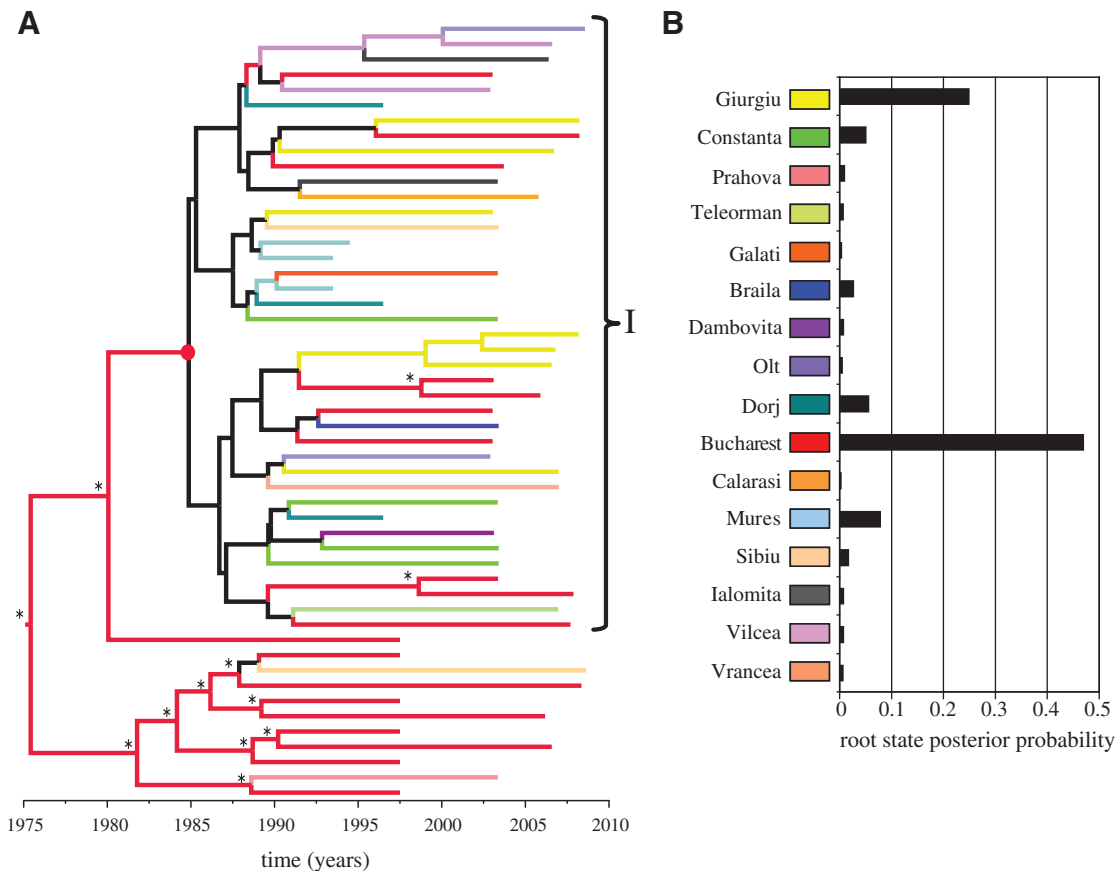


FIG. 4. Phylogeographic reconstruction of the Romanian F1 epidemic. **(A)** The MCC tree of all Romanian sequences with continuous-time Markov chain (CTMC) spatial reconstruction. Lineages are colored according to the highest posterior probability for the county of origin as shown in the legend in **B**; black represents an undetermined root state location. The asterisk on the MCC tree represents a root state probability > 0.8 for the county shown. **(B)** The bar graph shows the root state posterior probability for the parenteral epidemic ancestral node (red dot in **A**) for all the counties.

Second, we show that the parenteral epidemic has mostly maintained a constant population size after an initial exponential growth period that lasted up to the beginning of the 1990s. This suggests that very little parenteral transmissions and onward spread of the virus within this population has occurred since the early 1990s. This can be rationalized by the fact that the infections occurred in children who are unlikely to be engaged in high risk behavior and that public health measures to control the spread of the epidemic initiated in the early 1990s were a major success.

Lastly, phylogeographic analysis indicated that the root state location of the epidemic in the early 1980s was the adult population in Bucharest. However, it is possible that the location does not represent the ultimate source of the parenteral epidemic as the children could have relocated within Romania or worldwide since acquiring the infection, which occurred after the fall of the Communist regime in 1989. This could explain the presence of the three non-Romanian sequences within the monophyletic cluster from the United States, Germany, and Austria (Fig. 2). Further analysis with increased sampling of F1 sequences from both children and adult patients in the other significantly affected counties in Romania will need to be undertaken to obtain a clear phylogeographic picture of the Romanian parenteral epidemic.

Sequence Data

GenBank accession numbers of the subtype F1 HIV-1 sequences from EuroCoord-CHAIN generated and used in this study are JQ280885–JQ280942.

Acknowledgments

We especially thank Hannah Castro, David Dunn, Ali Judd, and the MRC Clinical Trials Unit for administrative and technical support for the EuroCoord-CHAIN study. We also thank all the members of the EuroCoord-CHAIN study group and all the patients, staff, and project management staff of the cohorts participating in the EuroCoord-CHAIN study, in particular those from the PENTA-EPPICC network. The research leading to these results has received part funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the project "Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN)," grant agreement no. 223131.

Author Disclosure Statement

No competing financial interests exist.

References

1. Patrascu IV, Constantinescu SN, and Dublanchet A: HIV-1 infection in Romanian children. *Lancet* 1990;335(8690):672.
2. Hersh BS, Popovici F, Apetrei RC, *et al.*: Acquired immunodeficiency syndrome in Romania. *Lancet* 1991;338(8768):645–649.
3. Hersh BS, Popovici F, Jezek Z, *et al.*: Risk factors for HIV infection among abandoned Romanian children. *AIDS* 1993;7(12):1617–1624.
4. Patrascu IV and Dumitrescu O: The epidemic of human immunodeficiency virus infection in Romanian children. *AIDS Res Hum Retroviruses* 1993;9(1):99–104.
5. Apetrei C, Necula A, Holm-Hansen C, *et al.*: HIV-1 diversity in Romania. *AIDS* 1998;12(9):1079–1085.
6. Hemelaar J, Gouws E, Ghys PD, and Osmanov S: Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006;20(16):W13–W23.
7. Apetrei C, Lousert-Ajaka I, Collin G, *et al.*: HIV type 1 subtype F sequences in Romanian children and adults. *AIDS Res Hum Retroviruses* 1997;13(4):363–365.
8. Dumitrescu O, Kalish ML, Kliks SC, Bandea CI, and Levy JA: Characterization of human immunodeficiency virus type 1 isolates from children in Romania: Identification of a new envelope subtype. *J Infect Dis* 1994;169(2):281–288.
9. Robertson DL, Anderson JP, Bradac JA, *et al.*: HIV-1 nomenclature proposal. *Science* 2000;288(5463):55–56.
10. Bandea CI, Ramos A, Pieniazek D, *et al.*: Epidemiologic and evolutionary relationships between Romanian and Brazilian HIV-subtype F strains. *Emerg Infect Dis* 1995;1(3):91–93.
11. Guimaraes ML, Vicente AC, Otsuki K, *et al.*: Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology* 2009;6:39.
12. Mehta SR, Wertheim JO, Delpont W, *et al.*: Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. *Infect Genet Evol* 2011;11(5):975–979.
13. Tamura K, Dudley J, Nei M, and Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24(8):1596–1599.
14. Stamatakis A, Ludwig T, and Meier H: RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005;21(4):456–463.
15. Guindon S and Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52(5):696–704.
16. Drummond AJ and Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.
17. Hue S, Pillay D, Clewley JP, and Pybus OG: Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci USA* 2005;102(12):4425–4429.
18. Lemey P, Rambaut A, Drummond AJ, and Suchard MA: Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009;5(9):e1000520.

Address correspondence to:

Jean L. Mbisa
 Antiviral Unit
 Virus Reference Department
 MS Colindale
 HPA
 61 Colindale Avenue
 London NW9 5EQ
 United Kingdom

E-mail: tamyo.mbisa@hpa.org.uk