BAYESIAN PHYLOGEOGRAPHIC ANALYSES CLARIFY THE ORIGIN OF THE HIV-1 SUBTYPE A VARIANT CIRCULATING IN FORMER SOVIET UNION'S COUNTRIES

Díez-Fuertes F, Cabello M, Thomson MM. Bayesian phylogeographic analyses clarify the origin of the HIV-1 subtype A variant circulating in former Soviet Union's countries. Infect Genet Evol. 2015 Jul;33:197-205. doi: 10.1016/j.meegid.2015.05.003. Epub 2015 May 5. PMID: 25952568

BAYESIAN PHYLOGEOGRAPHIC ANALYSES CLARIFY THE ORIGIN OF THE HIV-1 SUBTYPE A VARIANT CIRCULATING IN FORMER SOVIET UNION'S COUNTRIES

Francisco Díez-Fuertes[a], Marina Cabello[b,c], Michael M. Thomson[b,*]

[a] *AIDS Immunopathology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220 Majadahonda, Madrid, Spain*

[b] *HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220 Majadahonda, Madrid, Spain*

[c] *Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil*

**Corresponding author: Michael M. Thomson, e-mail: mthomson@isciii.es, telephone: +34 913223900*

Abstract

The HIV-1 subtype A variant dominating the HIV-1 epidemics in former Soviet Union (FSU) countries (A_{FSU}) represents one of the major clades of the HIV-1 pandemic. This variant was reported to have begun spreading among injecting drug users (IDUs) in the Ukrainian city of Odessa in late 1994. Two competing hypotheses have been proposed on the ancestral origin of the A_{FSU} variant, locating it either in the Democratic Republic of Congo (DRC) or in the Republic of Guinea (RG). The studies supporting these hypotheses employed phylogenetic analyses to identify HIV-1 sequences collected outside FSU countries ancestrally related to A_{FSU}. A different approach, based on Bayesian phylogenetic inference and coalescent-based population genetics, has been employed here to elucidate the ancestry of this HIV-1 variant and to improve our knowledge on its spread in FSU countries. The analyses were carried out using *env* (C2-V3-C3) and $p24^{gag}$ fragments of the HIV-1 genome. The inferred migration for the HIV-1 A_{FSU} variant revealed only one significantly supported migration pathway from Africa to Eastern Europe, supporting the hypothesis of its origin in the DRC and estimating the upper limit of the migration of the ancestral virus from Africa around 1970. The support for an origin in the RG was negligible. The results supported the main role of Odessa as the epicenter of the A_{FSU} epidemic, dating the tMRCA of the A_{FSU} variant around 1984, ten years before its explosive expansion among IDUs. The estimated origin of the A_{FSU} subcluster responsible for the IDU outbreak was also located in Odessa, with the estimated tMRCA around 1993. Statistically supported migration routes from Odessa to other cities of Ukraine, Russia, Kazakhstan, Uzbekistan and Belarus were also inferred by the Bayesian phylogeographic analysis. These results shed new light on the origin and spread of the HIV-1 A_{FSU} variant.

1. Introduction

In Eastern European and Central Asian countries formerly part of the Soviet Union, there was a 2.5-fold increase in the number of people living with HIV from 2001 to 2010, reaching an estimate of 1,300,000 in 2012 (UNAIDS Report, 2012). This was the greatest regional increase in the HIV-1 pandemic during this period. The Russian Federation and Ukraine represent almost 90% of HIV-1 infections in this region and have the highest number of people living with HIV relative to the general population, with prevalences in the adult population of 0.9% in Ukraine (UNAIDS Report, 2013) and between 0.8% and 1.4% in the Russian Federation (UNAIDS Report, 2012). The epidemic in this region is mainly concentrated among injecting drug users (IDUs)with approximately 40% of all new HIV infections in Eastern Europe and Central Asia being still transmitted in this population and percentages of IDUs living with HIV reaching the dramatic numbers of 52% in Estonia and 22% in Ukraine (UNAIDS Report, 2012).

Four phylogenetic lineages of HIV-1, groups M, N, O and P, have been described, which derive from independent cross-species transmissions of simian immunodeficiency viruses (SIVs) from African apes to humans. The M and N groups emerged after cross-species transmissions of SIV variants identified in chimpanzees of the *Pan troglodytes troglodytes* subspecies to humans (Keele et al., 2006), whereas groups O and P comprise viruses adapted to humans closely related to SIVs found in gorillas of the *Gorilla gorilla gorilla* subspecies (Plantier et al., 2009; D'arc et al., 2015). The evolution of these zoonotic events followed different paths, since groups N and P have only been found in a small number of individuals from Cameroon, whereas group O has spread epidemically in west central Africa and group M is the cause of the AIDS pandemic. Recently, group M has been estimated to have originated around 1920 in Kinshasa, capital of the Democratic Republic of Congo (DRC), remaining in a pre-epidemic state until its exponential growth, inferred around 1960 (Faria et al., 2014). HIV-1 group M comprises 9 subtypes and numerous recombinant forms which have spread around the world. In the case of the former Soviet Union's (FSU) countries, the HIV-1 epidemic is dominated by a subtype A variant (A_{FSU}), in contrast to the subtype B predominance in Western Europe (Bobkova, 2013). The epidemiological information indicates that A_{FSU} began to spread in Ukraine among IDUs in late 1994 (Hamers et al., 1997). In successive years, A_{FSU} outbreaks

among IDUs were also reported in Russia (Bobkov et al., 2001, 2004a), Belarus (Lukashov et al., 1998), and subsequently in most of the FSU countries, including the Baltic states - Estonia, Latvia and Lithuania (Balode et al., 2004; Caplinskas et al., 2013; Zetterberg et al., 2004), the Caucasus states - Georgia and Azerbaijan (Saad et al., 2006; Vinogradova et al., 2010), Central Asian states - Kazakhstan, Uzbekistan and Tajikistan (Beyrer et al., 2009; Bobkov et al., 2004b; Kurbanov et al., 2003) and Moldova (Pandrea et al., 2001). The A_{FSU} variant has rapidly become the most prevalent HIV-1 genetic form in all these states, with the exception of Estonia, where CRF06_cpx is more prevalent (Avi et al., 2014). Although the first cases of the A_{FSU} outbreak among IDUs were reported in late 1994 in Southern Ukraine, a retrospective analysis in the city of Odessa with samples collected in 1993 identified several cases phylogenetically related to the variant spreading among IDUs and predominantly transmitted through heterosexual contact, but with lower HIV-1 infection rates (Novitsky et al., 1998; Thomson et al., 2007).

The A_{FSU} represents one of the major clades identified in the global HIV-1 pandemic, leading to one of the most rapidly expanding HIV epidemics in the world. The ancestral origin of A_{FSU} variant remains a controversial issue due to the contradictory hypotheses that have been proposed. Based on phylogenetic analyses identifying sequences from outside FSU countries related to the A_{FSU} variant, Thomson et al. (2007) found several partial sequences from different segments of the HIV genome from viruses collected in the DRC (i.e., the *env* C2-V3-C3 region of KFE326 and 01CD106, a $p24^{gag}$ segment of KFE326, and integrase fragment of 01CD106) which branched basally to the A_{FSU} cluster. One year later, Riva et al. (2008) reported the identification of a virus characterized in the near full-length genome sequence, named 60000, collected in Italy from an individual presumably infected in the Republic of Guinea (RG), which was also related to A_{FSU}. The aim of the present study is to complement these phylogenetic studies with Bayesian phylodynamic approaches based on the Coalescence Theory to infer the most probable evolutionary hypotheses on the origin of A_{FSU} and computing the statistical support for both proposed hypotheses, locating the origin of A_{FSU} in Central or West Africa, respectively. Finally, the most probable routes of spread to and within the FSU countries of the HIV-1 A_{FSU} variant were also inferred by using a Bayesian phylogeographic approach.

2. Material and Methods

2.1. Datasets and phylogenetic inference

Three viruses have been proposed to be ancestrally related to A_{FSU}: KFE326 and 01CD106, collected in 1997 and 2001, respectively, in the DRC (Thomson et al., 2007), and 60000, collected in 2002 in Italy from a recent immigrant from the RG, where the virus would have been presumably acquired (Riva et al., 2008). The DRC viruses were identified as related to A_{FSU} through BLAST searches in database sequences followed by phylogenetic analyses of sequences most similar to A_{FSU} references (Thomson et al., 2007). Since these analyses were done eight years ago, in order to determine whether among the currently available sequences there were any other ancestrally related to the A_{FSU} variant, we downloaded all subsubtype A1 viruses from Africa and FSU from the databases (Los Alamos HIV Sequence Database) spanning the *env* fragment contained in KFE326 and 01CD106 sequences (positions 7056-7520 in HXB2 genome). The total number of sequences obtained was 6226 (5444 from Africa and 782 from FSU) and a filtering step was carried out in order to include only one sequence per individual using Bioedit (Hall, Ibis Biosciences). The genetic form of these sequences was checked using COMET tool (Struck et al., 1014), discarding 326 non-subsubtype A1 sequences erroneously assigned in the Los Alamos database and leaving 867 sequences for subsequent analyses. A multiple sequence alignment was carried out with an iterative refinement algorithm implemented in MAFFT (Katoh et al., 2005). The phylogeny of this dataset was inferred by an approximately maximum likelihood (ML) method using FastTree software version 2.1.7 (Price et al., 2009, 2010) and employing the general time reversible (GTR) substitution model with CAT approximation for among-site rate heterogeneity. A non-parametric branch support test based on a Shimodaira-Hasegawa-like (SH-like) procedure was employed (Shimodaira et al., 1999). The trees were visualized using FigTree software (FigTree v1.4.0 program).

2.2. Bayesian phylogeographic analyses

The optimal evolutionary models were selected based on the likelihood scores obtained by jModeltest v2.1.5 (Darriba et al., 2012) according to the Bayesian Information Criterion (BIC). The models selected for the *env* C2-V3-C3 and the $p24^{gag}$ datasets were the transversion model

with gamma-distributed rate variation among sites and a proportion of invariable sites (TVM+G+I) and HKY+G+I, respectively, which were applied to the analyses by manual modification of the xml files pre-generated by Beauti tool (Beast v1.8.0 software).. A subtype A dataset of the C2-V3-C3 fragment including 01CD106, KFE326 and 60000, as well as 77 other African sequences, was employed to infer the ancestral origin of the A_{FSU} cluster (Supplementary_file_1). Two other datasets based on the C2-V3-C3 (positions 6858-7647 in HXB2 genome) and $p24^{gag}$ (positions 1603-1998 in HXB2 genome) fragments corresponding to the sequences obtained from viruses collected in Odessa in 1993 and 1996 were also analysed to determine the origin of the cluster within the FSU. In order to avoid a sampling bias due to the overrepresentation of samples collected in Odessa, the analysis was repeated three times, discarding in a random manner subsets of sequences collected in this city and reducing the number of sequences from Odessa to 10 for the three replicates (15.4%), similar to the representatives of other cities such as Tashkent (12.3%) and about half of the sequences from Saint Petersburg (33.8%).

Several analyses were carried out with BEAST package in order to find the optimal combination of molecular clock model and coalescent tree priors that provide the best fit to the datasets (Drummond et al., 2012). Thus, the location was considered as a discrete trait employing a symmetric diffusion model (Wallace et al., 2007). The coalescent Gaussian Markov random field (GMRF) Bayesian Skyride tree prior (Minin et al., 2008) and an uncorrelated lognormal clock model were selected. The Markov Chain Monte Carlo (MCMC) was run for 100 million steps and convergence was checked using Tracer v1.5 by effective sample size (ESS) after a burn-in of 10 million steps. The analysis was run twice to reduce the risk of overfitting and each replicate was associated with a different value of the seed responsible for the random partitioning of the dataset. Finally, LogCombiner v1.7.5 was employed to combine the results obtained for each replicate. We report the evolutionary rate, the age of the most recent common ancestor and the demographic history of the A_{FSU} clade.

The existence of temporal signal in the data set employed to infer the ancestral origin of the A_{FSU} clade was assessed by a regression analysis of root-to-tip distances against dates of sampling. The correlation coefficient for the data set was calculated with Path-O-Gen v1.4 software (Rambaut et al., 2009) employing the trees inferred for the A_{FSU} cluster. Bayesian

uncertainty in parameter estimation was indicated by 95% highest posterior density (HPD) intervals.

2.3. Reconstructing the evolutionary history of the A_{FSU} variant

The results obtained by the Bayesian phylogeographic analysis were summarized using TreeAnnotator v1.7.5 with maximum clade credibility (MCC) genealogies, employing the mean value as the node height and discarding the first 10% of the generated trees as burn-in. The MCC genealogies were visualized with FigTree v1.4.0 program. The most probable migration routes and more significant rates of spread were also inferred using a Bayesian stochastic search variable selection (BSSVS) procedure (Lemey et al., 2009) and were visualized with the spatial phylogenetic reconstruction of evolutionary dynamics (SPREAD) tool (Bielejec et al., 2011). Statistical support for migration pathways was determined with Bayes factors (BF), and only routes supported by BFs over 10 were considered as strong evidence against the null hypothesis (Kass and Raftery, 1995).

3. Results

3.1. Phylogenetic analyses

The phylogenetic tree of the C2-V3-C3 fragment of all available subtype A sequences from FSU and Africa revealed a clade comprising FSU samples and the sequence 60000 (Fig. 1), identified in a 25-year-old male from Conakry, Republic of Guinea, who was diagnosed of HIV-1 infection in Perugia, Italy, in 2002, just two months after moving to Italy (Riva et al., 2008). Two African viruses were basally connected to this cluster, 01CD106 and KFE326 (both collected in the DRC) and no additional African sequences were phylogenetically related to A_{FSU} (Fig. 1).

The A_{FSU} IDU subcluster was supported by an SH-like value of 0.98 (Fig. 1), and included sequences from Russia, Ukraine, Armenia, Uzbekistan, Kazakhstan, Georgia and Belarus. Several sequences exhibited basal branching relative to this subcluster, all but one from viruses collected in the FSU, along with 60000. All sequences basally connected to the A_{FSU} IDU cluster (including 60000) were from individuals who denied intravenous drug use and

were reportedly infected via heterosexual contact and were from samples mainly collected in the city of Odessa (Novitsky et al., 1998).

3.2. Estimation of clock-like structure and evolutionary rate in datasets

The analysis of root-to-tip correlations of sampling time versus genetic distance revealed a clock-like structure in the case of C2-V3-C3 and $p24^{gag}$ fragments of the viruses included in the A_{FSU} IDU subcluster (r^2=0.6435 and r^2=0.3787, respectively). In the analyses of the A_{FSU} IDU subcluster together with sequences branching basally to it, the results supported the temporal clock-like divergence of the C2-V3-C3 fragment (r^2=0.3941) and suggested a weaker correlation for the $p24^{gag}$ fragment (r^2=0.2711). Therefore, the datasets of the C2-V3-C3 fragment, and to a lesser extent of the $p24^{gag}$ fragment, contained sufficient temporal structure for reliable estimation of evolutionary rates and divergence times of the A_{FSU} cluster (Supplementary_file_2). Finally, the rates of nucleotide substitution per site per year of the A_{FSU} cluster were calculated with BEAST, obtaining values of 6.94×10^{-3} (95% HPD, 5.41×10^{-3} – 8.88×10^{-3}) in the C2-V3-C3 fragment and 3.59×10^{-3} (95% HPD, 2.52×10^{-3} – 4.77×10^{-3}) in the $p24^{gag}$ fragment.

3.3. African ancestry of the A_{FSU} variant

Both proposed hypotheses on the origin of the A_{FSU} variant were statistically tested through the inferred MCC genealogy. Thus, the analysis of the C2-V3-C3 region placed 60000 virus within the A_{FSU} clade and KFE326 and 01CD106 branching basally to it (Fig. 4). The statistical support for the hypothesis placing the origin of the A_{FSU} variant in the RG was extremely low [posterior probability (PP) = 0.0003]. KFE326 was the African virus most closely related to A_{FSU} in the MCC genealogy, supporting the DRC as its most probable ancestral location (PP=0.84) with an estimated tMRCA of the clade comprising KFE326 and A_{FSU} at 1970 (95% HPD interval, 1962-1976), which represents the upper limit of migration from the DRC to Eastern Europe. The next most probable location at this node was Odessa, supported by a PP about six times lower (PP=0.13) than that of the DRC. Finally, 01CD106 was basally connected

to both A_{FSU} and KFE326. The most probable location of the origin at the node of the clade comprising 01CD106, KFE326 and A_{FSU} was the DRC (PP=0.94) with an estimated tMRCA at 1967 (95% HPD interval, 1958-1974) (Fig. 2).

3.4. Most probable migration pathways from Africa to the FSU

The BSSVS analysis of the C2-V3-C3 fragment detected two significant migration routes between Africa and Eastern Europe. The most probable migration route was between the DRC and Odessa supported by a BF of 70.1. The second was between RG and Odessa, supported by a BF of 18.5. However, the BSSVS analysis together with the temporally- and geographically-annotated MCC tree support the introduction of the ancestor of A_{FSU} into Eastern Europe from the DRC, with the route between RG and Odessa having the opposite directionality, placing Odessa as the most probable origin and the RG as the most probable destination.

3.5. Origin of the A_{FSU} cluster in FSU countries

The phylogeographic analysis of the $p24^{gag}$ segment revealed that the A_{FSU} variant originated in Odessa (PP=0.69) around 1984 (95% HPD interval, 1982-1987), approximately a decade before its expansion among IDUs (Fig. 3). The initial propagation of A_{FSU} was most likely driven by heterosexual transmission, since the most basally-branching viruses correspond to individuals infected via heterosexual contact not later than 1993 (Novitsky et al., 1998; Thomson et al., 2007). The analysis of the C2-V3-C3 region also placed the origin of A_{FSU} in Odessa (PP=0.99) with a mean estimated tMRCA in 1984 (95% HPD, 1980-1987) (Fig. 4). The combination of both analyses resulted in an overlapped credibility interval for the tMRCA of A_{FSU} of 1982-1987.

The origin of the A_{FSU} IDU subcluster was also tested with the Bayesian approach described above. The separate analyses carried out with the $p24^{gag}$ and C2-V3-C3 segments placed its most recent common ancestor in Odessa (PP=1 in both cases) with tMRCA in 1993

(95% HPD, 1992-1994) for both $p24^{gag}$ and C2-V3-C3 (Fig. 3 and 4). In order to examine whether the overrepresentation of samples from Odessa could have affected the results, three replicates of the C2-V3-C3 dataset were analyzed after the random removal of samples from Odessa to equalize them with samples from other locations, such as Saint Petersburg or Tashkent. The results obtained for the three replicate datasets were consistent with the previous analysis, confirming Odessa as the origin of the A_{FSU} cluster and the IDU subcluster (Supplementary_file_3).

3.6. Migration pathways of the A_{FSU} variant within FSU countries

The BSSVS analysis of the C2-V3-C3 segment revealed 11 strongly supported diffusion pathways between 13 different cities of Ukraine, Russia, Belarus, Uzbekistan and Kazakhstan (Table 1). The most strongly supported pathways were between Odessa and the cities of Shymkent (Kazakhstan), Kiev (Ukraine), Donetsk (Ukraine) and Tashkent (Uzbekistan), all with BFs above 180. Four other pathways connecting Odessa to Karaganda (Kazakhstan), Irkutsk (Russia), Poltava (Ukraine) and Svetlogorsk (Belarus) were supported by BFs > 10. All these pathways were also strongly supported in the analysis of the $p24^{gag}$ fragment, with the exception of those connecting Odessa to Donetsk and to Irkutsk (Table 1). Seven other pathways between FSU cities were identified with a weaker support ($3<BF<10$) in the $p24^{gag}$ analysis, connecting Odessa to Pavlodar (Kazakhstan) (BF=6.2) and to St. Petersburg (Russia) (BF=4.9), St. Petersburg to Novosibirsk (Russia) (BF=3.1) and to Samara (Russia) (BF=3.5), Shymkent to Pavlodar (Kazakhstan) (BF=8.8) and to Samara (BF=7.3) and Poltava to Svetlogorsk (BF=7.6).

4. Discussion

A_{FSU} is one of the major clades of the HIV-1 pandemic and a better knowledge of the origin and spread of this variant may be of public health relevance in FSU countries. The current information is contradictory, with different studies favoring two alternative hypotheses as the most plausible explanations for the origin of the epidemic. Both employed phylogenetic analyses to determine ancestral relationships to HIV-1 variants collected outside FSU countries. However, the increased knowledge of molecular evolution allowed us to employ more adequate mathematical methods to assess the geographical origin of a determined clade of phylogenetically related sequences. The main limitation for all these analyses, including the

phylogenetic analyses, is probably the capacity to obtain a subset of well-characterized samples able to represent the HIV-1 epidemic as real as possible in terms of sequence diversity and reliable epidemiologic information. In the present study, all the subtype A HIV-1 variants spanning the C2-V3-C3 fragment and collected in Africa and FSU countries were obtained from databases and only a few samples were discarded because of lack of epidemiological information.

One of the hypotheses places the ancestry of A_{FSU} in the DRC (Thomson et al., 2007) and the other in the RG (Riva et al., 2008). The analyses employing phylogeographic methods carried out in the present study support the origin of this HIV-1 variant in the DRC, estimating the upper limit of the migration of the African ancestor of A_{FSU} to Eastern Europe around 1970. According to our analyses, the support for the hypothesis placing the ancestry of A_{FSU} in the RG is negligible (PP=0.0003). The approximately maximum likelihood phylogenetic analyses showed that 60000 (i.e., the virus proposed to support the hypothesis of the origin of A_{FSU} in the RG) branched basally within the A_{FSU} clade, together with heterosexually-transmitted viruses collected in Odessa and St. Petersburg, a topological feature which was also observed in the Bayesian MCC tree. The BSSVS analysis showed a significantly supported migration between Odessa and the RG (BF=18.5), with the MCC tree pointing to Odessa as the origin and the RG as the destination of this migration. Taken together, the results suggest that 60000 rather than the African strain with the closest ancestral relation to the A_{FSU} variant, most likely derives from the initial A_{FSU} radiation in Southern Ukraine which spread via heterosexual contact.

We are unaware of any particular historical event which could have favored the migration of an HIV-1 variant from the DRC to Ukraine during the period estimated for that of the A_{FSU} ancestor. This could have been a chance phenomenon, similarly to the exportation outside of Africa and subsequent propagation of other HIV-1 variants, such as CRF06_cpx in Estonia (Zetterberg et al., 2004), of CRF02_AG in Uzbekistan (Carr et al., 2005) or of CRF01_AE in Thailand (Gao et al., 1996). However, we note that one of the earliest cases of HIV-1 transmission in the FSU, dated in 1982, corresponds to a mother-to-child transmission of a subtype F virus from a woman heterosexually-infected from a DRC man in St. Petersburg, with this virus originating a variant with a limited local propagation via heterosexual contact in this city (Fernández-García et al., 2009).

Novitsky et al. (1996) reported one of the first molecular epidemiological studies on the A_{FSU} variant identifying phylogenetically-related viruses collected in Odessa in 1993 from individuals reportedly infected by heterosexual contact, and in 1996 mainly from IDUs. In the 1990s, the port of Odessa, located on the northwestern shore of the Black Sea, was an important commercial port and transportation hub in the region. The inclusion of the samples described by Novitsky et al. (1996) in our analyses was key to discern the origin of the A_{FSU} clade within the FSU. According to the epidemiological information and phylogenetic analyses, the A_{FSU} variant was circulating via heterosexual contacts at relatively low transmission rates in the city of Odessa several years before its explosive spread among IDUs (Novistky et al., 1998, Thomson et al., 2007). This is in accordance with the results obtained in the present study, which dated the tMRCA of this cluster around 10 years before the first reported cases of HIV-1 transmissions among IDUs. In this context, the port of Odessa became the most important seaport for trade of the Soviet Union as well as a Soviet naval base, favoring a rapid growth of the city due to the migration of rural population and the attraction of industrial professionals from across the Soviet Union (Polmar, 1986). According to the results, the A_{FSU} variant initially found in heterosexuals in the city of Odessa with low transmission rates was introduced among IDUs probably between 1992 and 1994, which is consistent with the epidemiological information, which reported the first cases of transmission among IDUs in late 1994, initiating an explosive spread in this population, concomitantly with the deterioration of political and socioeconomic conditions of the Ukrainian population after the disintegration of the Soviet Union in 1991. The increase in drug abuse and the establishment of the new state of Ukraine as a stopping place in the newly emerging drug trafficking routes departing mainly from producing countries in Central Asia such as Afghanistan provided the perfect scenario for the propagation of this variant (Rhodes et al., 1999).

In successive years, most of the FSU countries reported HIV-1 epidemics caused by viruses of the A_{FSU} variant (Lukashov et al., 1998; Bobkov et al., 2001, 2004a, 2004b; Pandrea et al., 2001; Kurbanov et al., 2003; Balode et al., 2004; Zetterberg et al., 2004; Saad et al., 2006; Beyrer et al., 2009; Vinogradova et al., 2010; Caplinskas et al., 2013). The present study finds different routes of spread of this variant according to the inferred MCC genealogy and the BSSVS analysis. The city of Odessa is involved in eight of the 11 routes identified, confirming

its role as a local center of spread in the epidemic. Statistically significant routes of spread were identified connecting Odessa to other Ukranian cities (Kiev, Donetsk and Poltava) and to other cities of the FSU countries, such as Shymkent and Karaganda (Kazakhstan), Tashkent (Uzbekistan) and Irkutsk (Russia). The analysis of the C2-V3-C3 fragment also revealed significantly supported rates involving two of the largest cities of FSU, St. Petersburg and Moscow, and revealing a plausible route of migration between Pavlodar and St. Petersburg and between the latter and Moscow. The migration of the A_{FSU} variant between Odessa and Pavlodar had a weaker support in this analysis. Taken together, these results seem to indicate that the condition of Odessa as a transportation hub and a major seaport of the Black Sea provided a perfect breeding ground for an explosive expansion of the A_{FSU} variant among IDUs in FSU countries.

A comment should be made on the evolutionary rates inferred by BEAST for the C2-V3-C3 fragment (6.94×10^{-3} substitutions site^{-1} year^{-1}), which was higher than those previously obtained for the whole envelope gene for group M viruses (3.26×10^{-3} substitutions site^{-1} year^{-1}) (Faria et al., 2014) and for the A_{FSU} IDU clade using sequences spanning 260 to 300 bp of the V3 *env* region (2.50×10^{-3} substitutions site^{-1} year^{-1}) (Maljkovic Berry et al., 2007). However, this apparent discrepancy could be explained by the analysis of different fragments, since the present study analyzed longer sequences, spanning positions 6858 to 7647 of HXB2 genome. Moreover, the studies analyzed viruses propagating in different populations, since the analysis carried out by Maljkovic Berry et al. (2007) was done only with viruses transmitted among IDUs, and our analyses also included heterosexually-transmitted viruses circulating in Ukraine. Regarding the evolutionary rate for the *gag* fragment, a value of 3.59×10^{-3} substitutions site^{-1} year^{-1} was obtained which is higher than those obtained for subtype B viruses (1.1×10^{-3} substitutions site^{-1} year^{-1}) (Alizon et al., 2013). However, the analysis of CRF01_AE viruses, whose *gag* gene corresponds to subtype A, revealed an evolutionary rate of 2.4×10^{-3} substitutions site^{-1} year^{-1} (Tee et al., 2009), which is closer to that obtained in the present work.

5. Conclusions

In summary, the results of the Bayesian phylogeographic analyses strongly support an ancestry of the A_{FSU} variant in the DRC and its initial spread in Odessa via heterosexual transmission about a decade before its explosive propagation through a single point introduction among IDUs. The support for the alternative hypothesis proposing an ancestry of A_{FSU} in the RG (Riva et al., 2008) was extremely low; rather, the results suggested that the virus proposed to represent the ancestral variant of A_{FSU} might derive from the initial A_{FSU} radiation in Odessa. The methods employed in the present study were validated by the temporal and geographic coincidence of the known epidemiological data of the IDU outbreak and the estimations obtained by the phylogeographical analyses. These results provide important epidemiological insights on the origin and spread of one of the major regional HIV-1 epidemics affecting a vast geographic area and a large number of individuals.

6. Acknowledgments

Francisco Díez-Fuertes is supported by the Sara Borrell postdoctoral Program of the Spanish Government 2012 CD12/00515.

7. References

Alizon, S., Fraser, C., 2014. Within-host and between-host evolutionary rates across the HIV-1 genome. Retrovirology 10: 49.

Avi, R., Huik, K., Pauskar, M., Ustina, V., Karki, T., Kallas, E., Jõgeda, E.L., Krispin, T., Lutsar, I., 2014. Transmitted drug resistance is still low in newly diagnosed human immunodeficiency virus type 1 CRF06_cpx-infected patients in Estonia in 2010. AIDS Res. Hum. Retroviruses 30, 278-283.

Balode, D., Ferdats, A., Dievberna, I., Viksna, L., Rozentale, B., Kolupajeva, T., Konicheva, V., Leitner, T., 2004. Rapid epidemic spread of HIV type 1 subtype A1 among intravenous drug users in Latvia and slower spread of subtype B among other risk groups. AIDS Res. Hum. Retroviruses 20,245-249.

Beyrer, C., Patel, Z., Stachowiak, J.A., Tishkova, F.K., Stibich, M.A., Eyzaguirre, L.M., Carr, J.K., Mogilnii, V., Peryshkina, A., Latypov, A., Strathdee, S.A., 2009. Characterization of the

emerging HIV type 1 and HCV epidemics among injecting drug users in Dushanbe, Tajikistan. AIDS Res. Hum. Retroviruses 25, 853-860.

Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910-2912.

Bobkov, A., Kazennova, E., Khanina, T., Bobkova, M., Selimova, L., Kravchenko, A., Pokrovsky, V., Weber, J., 2001. An HIV type 1 subtype A strain of low genetic diversity continues to spread among injecting drug users in Russia: study of the new local outbreaks in Moscow and Irkutsk. AIDS Res. Hum. Retroviruses 17, 257-261.

Bobkov, A.F., Kazennova, E.V., Selimova, L.M., Khanina, T.A., Ryabov, G.S., Bobkova, M.R., Sukhanova, A.L., Kravchenko, A.V., Ladnaya, N.N., Weber, J.N., Pokrovsky, V.V., 2004a. Temporal trends in the HIV-1 epidemic in Russia: predominance of subtype A. J. Med. Virol. 74, 191-196.

Bobkov, A.F., Kazennova, E.V., Sukhanova, A.L., Bobkova, M.R., Pokrovsky, V.V., Zeman, V.V., Kovtunenko, N.G., Erasilova, I.B., 2004b. An HIV type 1 subtype A outbreak among injecting drug users in Kazakhstan. AIDS Res. Hum. Retroviruses 20, 1134-1136.

Bobkova, M., 2013. Current status of HIV-1 diversity and drug resistance monitoring in the former USSR. AIDS Rev. 15, 204-212.

Caplinskas, S., Loukachov, V.V., Gasich, E.L., Gilyazova, A.V., Caplinskiene, I., Lukashov, V.V., 2013. Distinct HIV type 1 strains in different risk groups and the absence of new infections by drug-resistant strains in Lithuania. AIDS Res. Hum. Retroviruses 29, 732-737.

Carr, J.K., Nadai, Y., Eyzaguirre, L., Saad, M.D., Khakimov, M.M., Yakubov, S.K., Birx, D.L., Graham, R.R., Wolfe, N.D., Earhart, K.C., Sanchez, J.L., 2005. Outbreak of a West African recombinant of HIV-1 in Tashkent, Uzbekistan. J Acquir Immune Defic Syndr. 39: 570-575.

D'arc, M., Ayouba, A., Esteban, A., Learn, G.H., Boué, V., Liegeois, F., Etienne, L., Tagg, N., Leendertz, F.H., Boesch, C., Madinda, N.F., Robbins, M.M., Gray, M., Cournil, A., Ooms, M., Letko, M., Simon, V.A., Sharp, P.M., Hahn, B.H., Delaporte, E., Mpoudi Ngole, E., Peeters, M.,

2015. Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc Natl Acad Sci U S A. In press.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969-1973.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O.G., Lemey, P. 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science 346, 56-61.

Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H., Sharp, P.M., 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J Virol. 70: 7013-7029.

Hamers, F.F., Batter, V., Downs, A.M., Alix, J., Cazein, F., Brunet, J.B., 1997. The HIV epidemic associated with injecting drug use in Europe: geographic and time trends. AIDS 11, 1365-1374.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773-795.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313, 523-526.

Kurbanov, F., Kondo, M., Tanaka, Y., Zalalieva, M., Giasova, G., Shima, T., Jounai, N., Yuldasheva, N., Ruzibakiev, R., Mizokami, M., Imai, M., 2003. Human immunodeficiency virus in Uzbekistan: epidemiological and genetic analyses. AIDS Res. Hum. Retroviruses 19, 731-738.

Lemey, P., Rambaut ,A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5, e1000520.

Lukashov, V.V., Karamov, E.V., Eremin, V.F., Titov, L.P., Goudsmit, J., 1998. Extreme founder effect in an HIV type 1 subtype A epidemic among drug users in Svetlogorsk, Belarus. AIDS Res. Hum. Retroviruses 14, 1299-1303.

Maljkovic Berry, I., Ribeiro, R., Kothari, M., Athreya, G., Daniels, M., Lee, H.Y., Bruno, W., Leitner, T., 2007. Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. J Virol 81: 10625-10635.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459-1471.

Novitsky, V.A., Montano, M.A., Essex, M., 1998. Molecular epidemiology of an HIV-1 subtype A subcluster among injection drug users in the Southern Ukraine. AIDS Res. Hum. Retroviruses 14, 1079-1085.

Pandrea, I., Descamps, D., Collin, G., Robertson, D.L., Damond, F., Dimitrienco, V., Gheorghita, S., Pecec, M., Simon, F., Brun-Vézinet, F., Apetrei, C., 2001. HIV type 1 genetic diversity and genotypic drug susceptibility in the Republic of Moldova. AIDS Res. Hum. Retroviruses 17, 1297-1304.

Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L., Simon, F. 2009. A new human immunodeficiency virus derived from gorillas. Nat Med. 15, 871-872.

Polmar, N., 1986. The Naval Institute guide to the Soviet Navy, fifth ed. United States Naval Institute, Maryland.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253-1256.

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641-1650.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Rhodes, T., Ball, A., Stimson, G.V., Kobyshcha, Y., Fitch, C., Pokrovsky, V., Bezruchenko-Novachuk, M., Burrows, D., Renton, A., Andrushchak, L., 1999. HIV infection associated with drug injecting in the newly independent states, eastern Europe: the social and economic context of epidemics. Addiction 94, 1323-1336.

Riva, C., Romano, L., Saladini, F., Lai, A., Carr, J.K., Francisci, D., Balotta, C., Zazzi, M., 2008. Identification of a possible ancestor of the subtype A1 HIV Type 1 variant circulating in the former Soviet Union. AIDS Res. Hum. Retroviruses 24, 1319-1325.

Saad, M.D., Aliev, Q., Botros, B.A., Carr, J.K., Gomatos, P.J., Nadai, Y., Michael, A.A., Nasibov, Z., Sanchez, J.L., Brix, D.I., Earhart, K.C., 2006. Genetic forms of HIV Type 1 in the former Soviet Union dominate the epidemic in Azerbaijan. AIDS Res. Hum. Retroviruses 22, 796-800.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Struck, D., Lawyer, G., Ternes, A.M., Schmit, J.C., Bercoff, D.P., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42, e144.

Tee, K.K., Pybus, O.G., Parker, J., Ng, K.P., Kamarulzaman, A., Takebe, Y., 2009. Estimating the date of origin of an HIV-1 circulating recombinant form. Virology 387: 229-234.

Thomson, M.M., de Parga, E.V., Vinogradova, A., Sierra, M., Yakovlev, A., Rakhmanova, A., Delgado, E., Casado, G., Muñoz, M., Carmona, R., Vega, Y., Pérez-Alvarez, L., Contreras, G., Medrano, L., Osmanov, S., Nájera, R., 2007. New insights into the origin of the HIV type 1 subtype A epidemic in former Soviet Union's countries derived from sequence analyses of preepidemically transmitted viruses. AIDS Res. Hum. Retroviruses 23, 1599-1604.

Vinogradova, A., Gafurova, E., Muñoz-Nieto, M., Rakhmanova, A., Osmanov, S., Thomson, M.M., 2010. Short communication: Molecular epidemiology of HIV type 1 in the Republic of Dagestan, Russian Federation: virtually uniform circulation of subtype A, former Soviet Union

variant, with predominance of the V77I(PR) subvariant. AIDS Res. Hum. Retroviruses 26, 395-400.

Wallace, R.G., Hodac, H., Lathrop, R.H., Fitch, W.M., 2007. A statistical phylogeography of influenza A H5N1. Proc Natl Acad Sci U S A 104, 4473-4478.

Zetterberg, V., Ustina, V., Liitsola, K., Zilmer, K., Kalikova, N., Sevastianova, K., Brummer-Korvenkontio, H., Leinikki, P., Salminen, M.O., 2004. Two viral strains and a possible novel recombinant are responsible for the explosive injecting drug use-associated HIV type 1 epidemic in Estonia. AIDS Res. Hum. Retroviruses 20, 1148-1156.

8. Web references

Hall, T. (Ibis Biosciences). Bioedit v.7.1.3.0: Biological sequence alignment editor. Available at: http://www.mbio.ncsu.edu/BioEdit/bioedit.html

FigTree v1.4.0. Available at: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A. Path-O-Gen: temporal signal investigation tool. Available at: http://tree.bio.ed.ac.uk/software/pathogen/

The Los Alamos HIV sequence database. Last accessed: 14 January, 2015. Available at: http://www.hiv.lanl.gov/

UNAIDS/WHO Report on the global AIDS epidemic 2012. Last accessed: 14 January, 2015. Available at: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_unaids_global_report_2012_with_annexes_en.pdf

UNAIDS/WHO Report on the global AIDS epidemic 2013. Last accessed: 14 January, 2015. Available at: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

Table 1. Significant non-zero rates for the Bayes factor test of A_{FSU} variant employing C2-V3-C3 and $p24^{gag}$ fragments. Only pathways supported by BFs ≥ 10 at any segment are shown.

Migration route		Bayes Factor	
		C2-V3-C3	**$p24^{gag}$**
Odessa	Shymkent	456683	212
Odessa	Kiev	231	10
Odessa	Donetsk	214	<10
Odessa	Tashkent	188	14
Odessa	Karaganda	29	12
Odessa	Irkutsk	24	<10
Odessa	Poltava	14	36
Odessa	Svetlogorsk	12	27
Pavlodar	St. Petersburg	59	<10
Moscow	Novosibirsk	56	<10
St. Petersburg	Moscow	19	<10

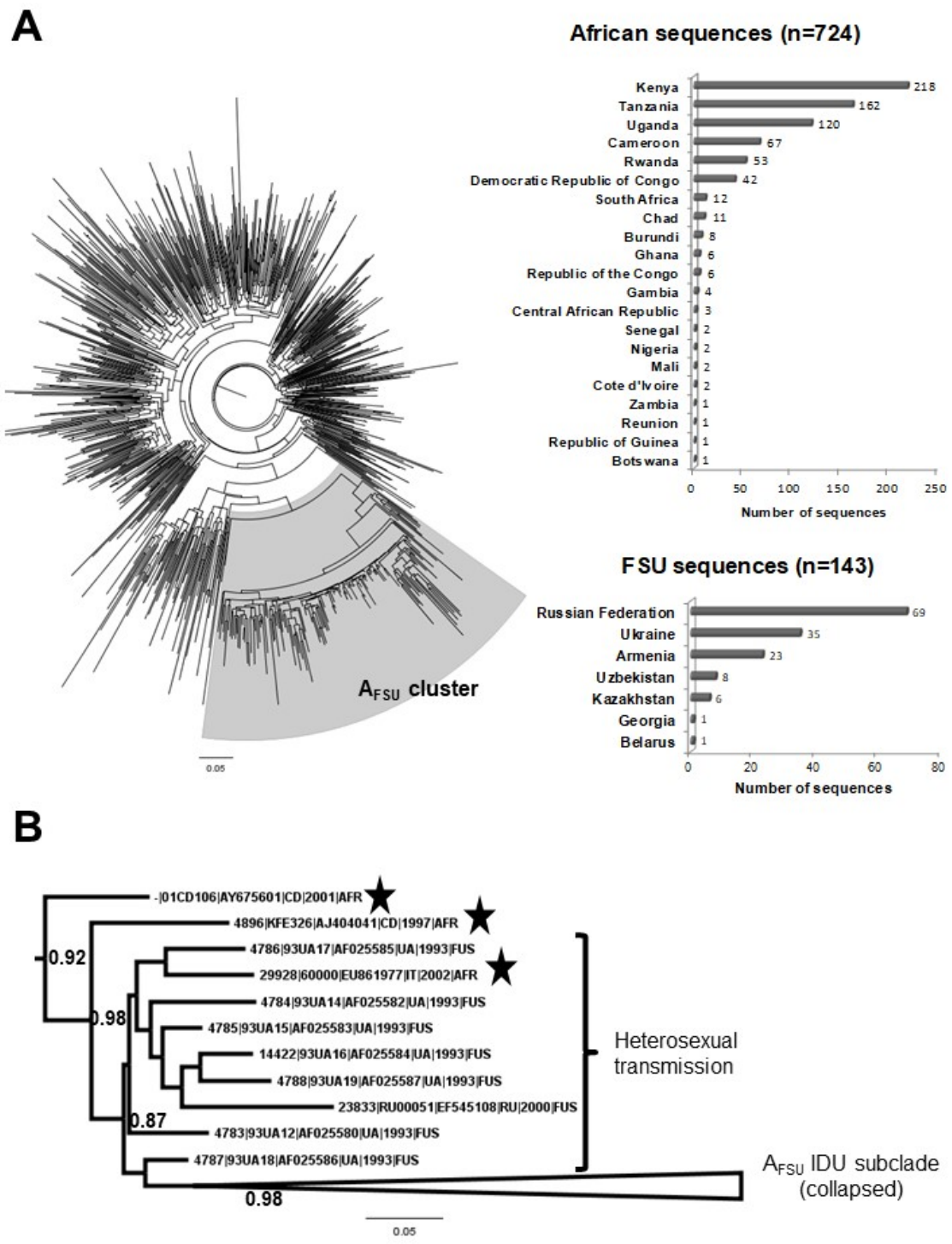


Figure 1. Approximately maximum likelihood phylogenetic analysis of a C2-V3-C3 fragment (positions 7056-7520 in HXB2 genome) including all the subsubtype A1 sequences available at the Los Alamos HIV Sequence Database from Africa and FSU countries. (A) Unrooted tree with highlighted A_{FSU} cluster and the African sequences branching basally to it, and graphs showing the geographic origin of all the samples included in the analysis. (B) Enlargement of the part of

the tree shown in (A) comprising the A_{FSU} cluster and the two African sequences branching basally to it. The sequences supporting both proposed hypotheses on the ancestral origin of the A_{FSU} cluster are marked with stars. SH-like support of relevant nodes are shown.

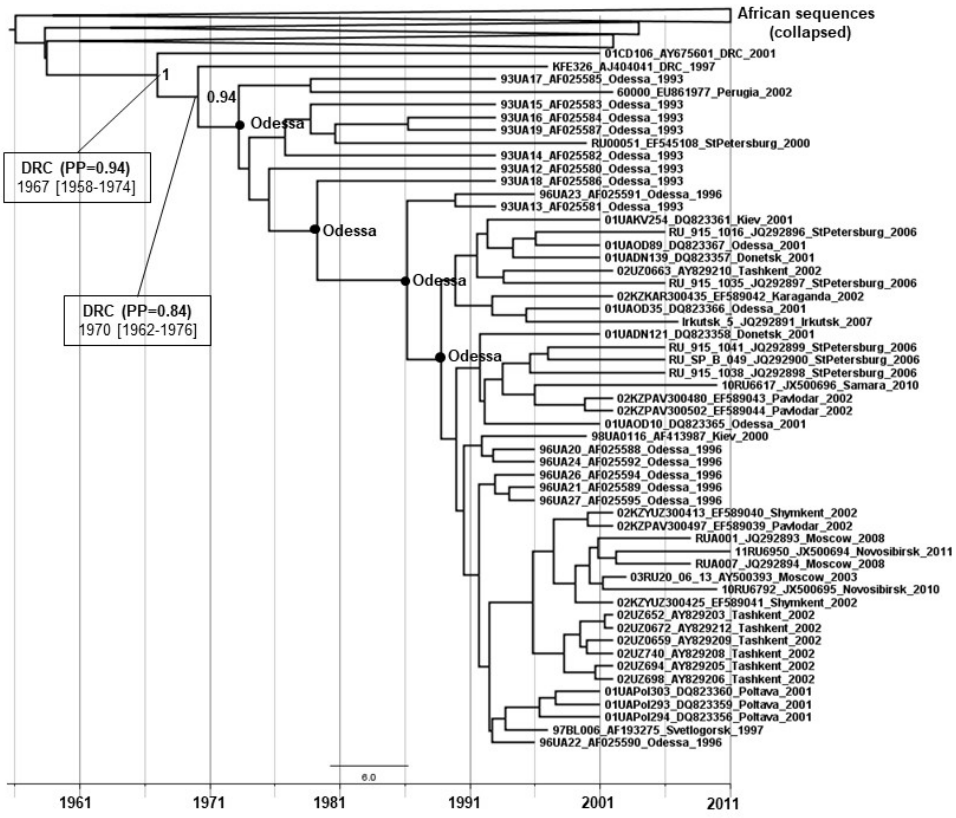


Figure 2. Maximum clade credibility genealogy of the C2-V3-C3 fragment of the A_{FSU} variant, together with the related viruses KFE326, 01CD106 and 60000. The nodes supported by PP ≥ 0.95 are marked with black circles, with indication of the place of the most probable location. For the nodes of clades comprising African sequences branching basally to A_{FSU}, the PP supporting the most probable location of the node and its estimated tMRCA with 95% HPD intervals is indicated. Name, accession number and city (or country in samples collected in DRC) and year of sample collection are indicated for each sample. All the African samples included in Supplementary_file_1 were included as outgroup for the A_{FSU} clade and shown compressed as triangles.

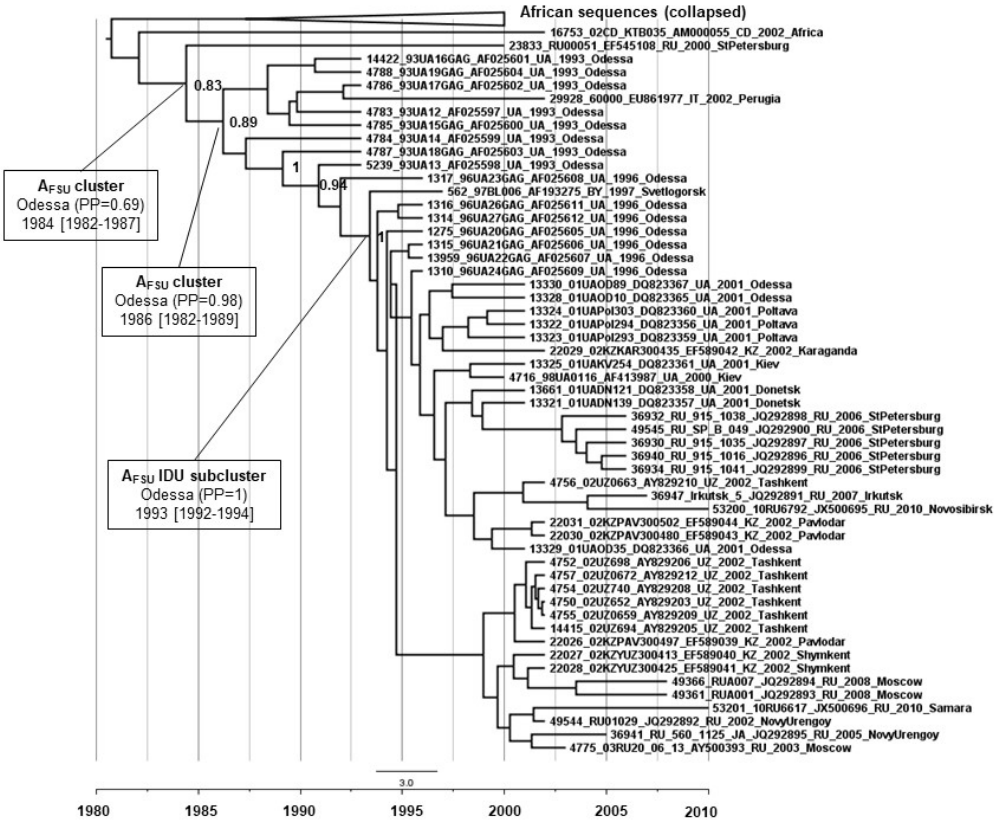


Figure 3. Maximum clade credibility genealogy of the $p24^{gag}$ fragment of the A_{FSU} cluster and the A_{FSU} IDU subcluster. The most probable city at the origin of the A_{FSU} variant and of the A_{FSU} IDU subcluster, together with location PP and tMRCA with 95% HPD intervals, is indicated. The patient code, the name and accession number of the virus, the ISO two-letter country code and the year and city of sample collection are indicated for each sample.

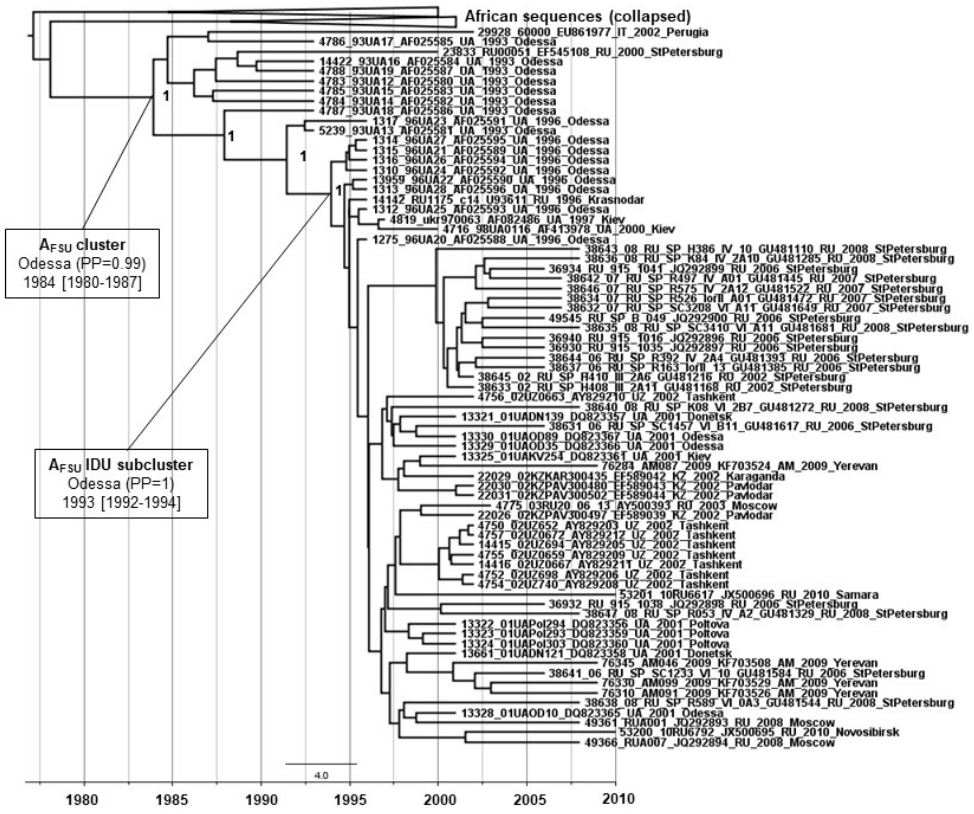
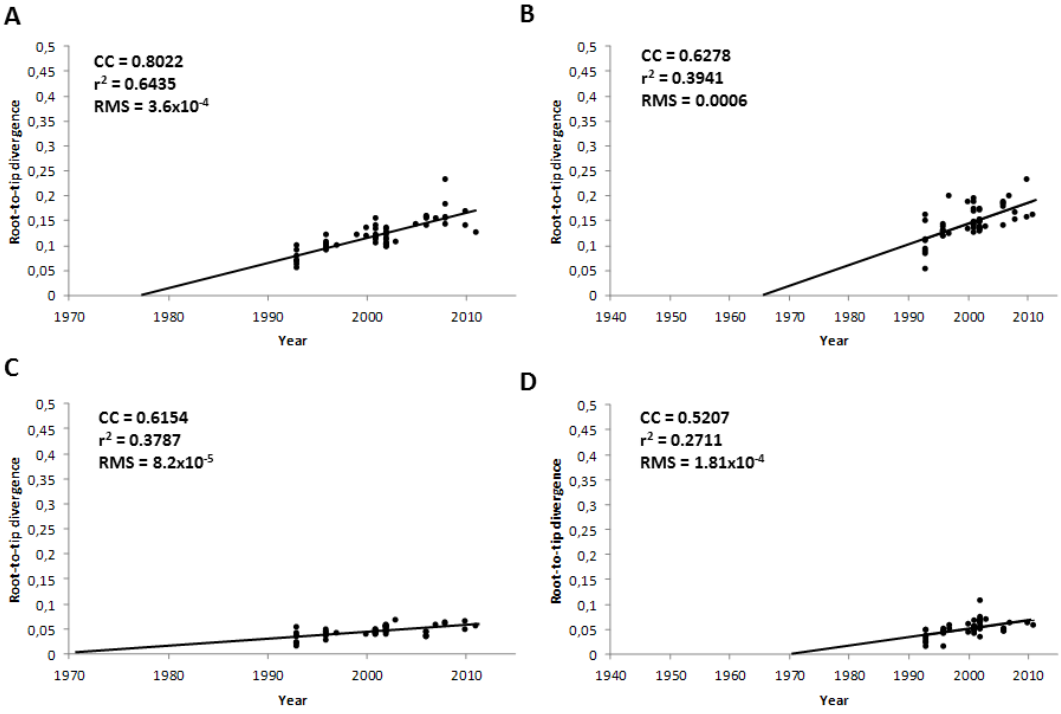
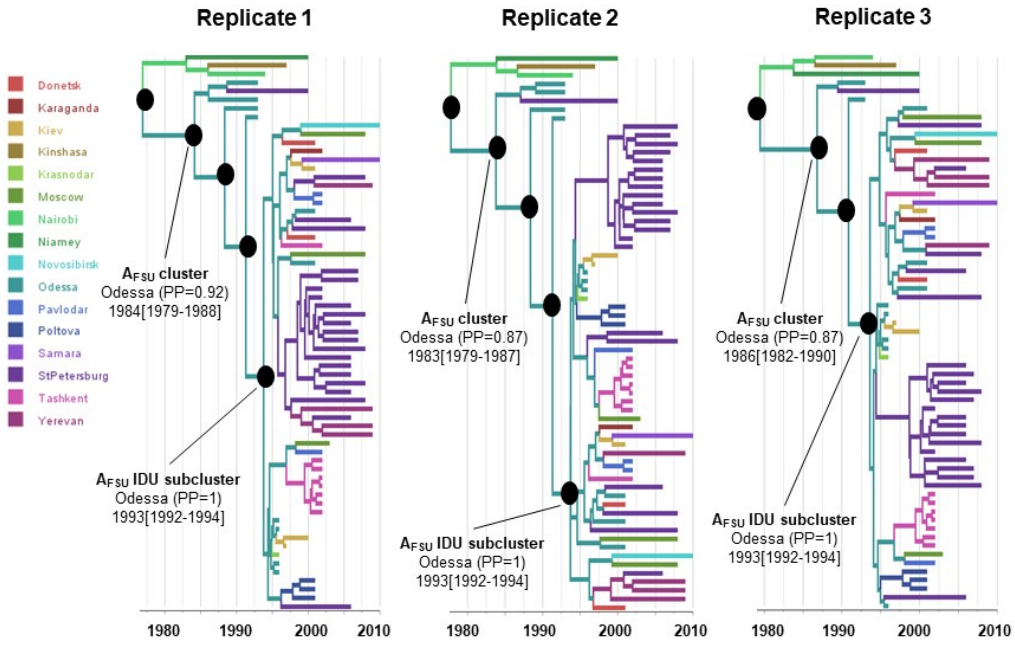


Figure 4. Maximum clade credibility genealogy of the C2-V3-C3 fragment of the A_{FSU} cluster and the A_{FSU} IDU subcluster. The most probable city at the origin of the A_{FSU} variant and of the A_{FSU} IDU subcluster, together with location PP and tMRCA with 95% HPD intervals, is indicated. The patient code, the name and accession number of the virus, the ISO two-letter country code and the year and city of sample collection are indicated for each sample.

Supplementary_file_1. African and FSU sequences employed for the reconstruction of ancestral location states of the A_{FSU} cluster. The temporal and geographic data associated to the sequences used in the analyses were, for FSU viruses, year, city and country of sample collection, and for African viruses, year and country of sample collection. The analyzed locus was 7056-7520 according to the HXB2 genome.



Supplementary_file_2. Analysis of the correlation of genetic distance with sampling year for different datasets: (A) C2-V3-C3 fragment of the A_{FSU} variant; (B) C2-V3-C3 fragment of the A_{FSU} variant together with KFE326 and 01CD106 of DRC; (C) $p24^{gag}$ fragment of the A_{FSU} variant; and (D) $p24^{gag}$ fragment of the A_{FSU} cluster together with KFE326. The r^2 estimates the fit of the data to a strict molecular clock by testing the degree of the influence that sampling time has over the amount of pairwise diversity in the data. CC = correlation coefficient; r^2 = regression coefficient; RMS = residual mean squared.



Supplementary_file_3. Maximum clade credibility genealogies of the C2-V3-C3 fragment of replicate subsets of viruses obtained after random removal of samples collected in Odessa. The most probable city at the origin of the A_{FSU} variant and of the A_{FSU} IDU subcluster, together with location PP and tMRCA with 95% HPD intervals, is shown. The nodes supported by PP ≥ 0.95 are marked with black circles.