Molecular epidemiology, phylogeny, and phylodynamics of CRF63_02A1, a recently originated HIV-1 circulating recombinant form spreading in Siberia

Nadezhda S Shcherbakova, Lyudmila A Shalamova, Elena Delgado, Aurora Fernández-García, Yolanda Vega, Larissa I Karpenko, Alexander A Ilyichev, Yuri V Sokolov, Dmitry N Shcherbakov, Lucía Pérez-Álvarez, Michael M Thomson.

AIDS Res Hum Retroviruses. 2014 Sep;30(9):912-9.

Title:

Molecular epidemiology, phylogeny and phylodynamics of CRF63_02A1, a recently originated HIV-1 circulating recombinant form spreading in Siberia.

Running head:

CRF63_02A1 phylogeny and phylodynamics

Authors:

Nadezhda S Shcherbakova[1]*, Lyudmila A Shalamova[2]*, Elena Delgado[3], Aurora Fernández-García[3], Yolanda Vega[3], Larissa I Karpenko[4], Alexander A Ilyichev[1], Yuri V Sokolov[5], Dmitry N Shcherbakov[6], Lucía Pérez-Álvarez[3], Michael M Thomson[3].

[1]Department of Bioengineering, State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk, Russia.

[2]Department of Zoonotic Infections and Influenza, State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk, Russia.

[3]HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain.

[4]Laboratory of Recombinant Vaccine, State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk, Russia.

[5]Center for Prevention and Control of AIDS and Infectious Diseases, Koltsovo, Novosibirsk, Russia.

[6]Altai State University, Barnaul, Russia.

*Both authors contributed equally to this work

Corresponding author:

Michael M Thomson, MD, PhD

Centro Nacional de Microbiología

Instituto de Salud Carlos III

Carretera Majadahonda-Pozuelo, Km. 2

28220 Majadahonda (Madrid)

Spain

Tel: +34-91-8223900

Fax: +34-91-5097014

e-mail: mthomson@isciii.es

Abstract

The HIV-1 epidemic in Russia is dominated by the former Soviet Union subtype A (A_{FSU}) variant but other genetic forms are circulating in the country. One is the recently described CRF63_02A1, derived from recombination between a CRF02_AG variant circulating in Central Asia and A_{FSU}, which has spread in the Novosibirsk region, Siberia. Here we phylogenetically analyze *pol* and *env* segments from 24 HIV-1 samples from the Novosibirsk region collected in 2013, with characterization of 3 new near full-length genome CRF63_02A1 sequences, and estimate the time of the most recent common ancestor (tMRCA) and the demographic growth of CRF63_02A1 using a Bayesian method. The analyses revealed that CRF63_02A1 is highly predominant in the Novosibirsk region (81.2% in *pol* sequences) and is transmitted both among injecting drug users and by heterosexual contact. Similarity searches with database sequences combined with phylogenetic analyses show that CRF63_02A1 is circulating in East Kazakhstan and the Eastern area of Russia bordering China. The analyses of near full-length genome sequences show that its mosaic structure is more complex than reported, with 18 breakpoints. The tMRCA of CRF63_02A1 was estimated around 2006, with exponential growth in 2008-2009 and subsequent stabilization. These results provide new insights into the molecular epidemiology, phylogeny and phylodynamics of CRF63_02A1.

Russia is one of the non-African countries with the greatest number of HIV-1-infected persons, only surpassed by India and the United States[1]. Before 1996 there were approximately 1,000 HIV-1-infected people in Russia (http://www.hivrussia.org/). In that period, subtype B was predominant, although other HIV-1 clades, of diverse geographic origins, were also detected[2-4]. In 1995-96 HIV-1 infections began to increase dramatically concomitantly with the expansion of a subtype A variant of Central African ancestry[5] that originated in Southern Ukraine[6,7] and spread to all countries of the former Soviet Union (FSU), in most of which it is the predominant HIV-1 genetic form[8,9], and for this reason it is frequently designated A_{FSU} variant. In addition to A_{FSU}, other HIV-1 genetic forms circulating in FSU countries at lower prevalences include subtype B, predominant in men who have sex with men[8-10], CRF03_AB, predominant in the Russian cities of Kaliningrad[11] and Cherepovets[12], subtype F, circulating as a minor variant in St. Petersburg, Russia[13], and a CRF02_AG variant ($CRF02\_AG_{FSU}$)[14,15], which was first detected among injecting drug users (IDUs) in Tashkent, Uzbekistan, in 1999-2000[14], and has subsequently been reported in Kazakhstan[15], Kyrgyzstan[16], and in the Novosibirsk region, Siberia, Russia[17]. This variant has generated, through secondary recombination with A_{FSU}, a new circulating recombinant form, CRF63_02A1, recently identified in Novosibirsk[17,18]. Here we examine the prevalence of CRF63_02A1 in the Novosibirsk region in samples collected in 2013, we obtain three new near full-length genome sequences of CRF63_02A1, reanalyzing its mosaic structure, and we estimate its epidemic history.

For this study, 26 serum samples from HIV-1-infected individuals were collected in May and June 2013 at the Center for Prevention and Control of AIDS and Infectious Diseases, Koltsovo, Novosibirsk region, which is located 5 km from the city of Novosibirsk and attends all HIV-infected people of the Novosibirsk region. This study was approved by the Ethics Committee of

the Center for Prevention and Control of AIDS and Infectious Diseases at Koltsovo, Novosibirsk region, Russia.

RNA was extracted from 1 ml plasma using Nuclisens EasyMAG kit (bioMérieux, Marcy l'Etoile, France) following the manufacturer's instructions.-The HIV-1 protease-reverse transcriptase (PR-RT) segment of *pol* and the C2-V3-C3 segment of *env* were amplified by RT-PCR followed by nested PCR. Primers used for PR-RT amplification were RP1-S and RP-1-A in RT-PCR and PR-O-S2b and RT-O-A in nested PCR (Table 1) and those used for amplification of the V3 region were described previously[19]. Near full-length genome (~9 kb) amplification in overlapping segments by RT-PCR and nested PCR and sequencing was done using a protocol similar to that described by us[20,21], although some amplified fragments and their corresponding primers, all of which are listed in Table 1, were different. For amplification of the 3' semigenome, primers SG3-up and SG3-lo were used for RT-PCR. This would allow for amplification of the full-length envelope sequence in nested PCR using primers Env-S and Env-A, which could be used for subsequent construction of functional envelope clones[22]. Both fragments flanking *env*, from *vif* through *vpu*, and *nef* plus a segment of the 3'LTR, respectively, were amplified by nested PCR from the RT-PCR product amplified with SG3-up and SG3-lo, using primers SG3-N-S and SSD2b, and NEF-S4 and 3'nef-3c, respectively. Gag and a fragment of the 5' LTR were amplified with SC-A-O-S-R and SC-A-O-A in RT-PCR and SC-A-N-S-R and SC-A-N-A in nested PCR; and the 3' segment of *pol* was amplified with RTDS-O-S and SC-B-N-A in RT-PCR and RTDS-N-S and B2-N-A in nested PCR. Sequence electropherograms were assembled with Seqman (DNASTAR, Madison, WI, USA).

Sequences were aligned with MAFFT v.7[23]. Phylogenetic trees were constructed via maximum likelihood with RAxML v.7.2.7[24], applying the general time reversible (GTR) substitution model with CAT approximation for among-site rate heterogeneity (GTR+CAT), with assessment of node support by bootstrapping. Recombination was analyzed by bootscanning

with Simplot v3.5[25], using a 200 nucleotide (nt) window with tree construction by the neighbor-joining method applying Kimura's two-parameter substitution model (the shorter than usual window size employed in the analysis was intended for a more precise definition of recombinant structures, with detection of fragments delimited by relatively close breakpoints). Short putatively recombinant segments (≤200 nt) identified with Simplot were further phylogenetically analyzed via maximum likelihood (ML) using RAxML, applying the GTR+CAT model, and with PhyML v3.0[26], applying the GTR with gamma-distributed rate heterogeneity among sites and a proportion of invariable sites (GTR+G+I) substitution model, with assessment of node support by the approximate likelihood ratio test (aLRT) using a Shimodaira-Hasegawa (SH)-like procedure. In this analysis, trees were constructed with reference sequences of CRF02_AG$_{FSU}$ and A$_{FSU}$ and were rooted with an inferred group M ancestral sequence downloaded from the HIV Sequence Database[27]. Segments were assigned to either variant based on clustering with a bootstrap value ≥70% with RAxML or an aLRT-SH-like support ≥0.8 with PhyML.

To identify CRF63_02A1 viruses from other geographical areas, we downloaded all HIV-1 sequences from FSU countries deposited at the HIV Sequence Database[27]. Similarity of these sequences to all available CRF63_02A1 and CRF02_AG near full-length genome sequences was examined with local BLAST searches using BioEdit v.7.1.3.0 (Tom Hall, www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequences most similar to two or more CRF63_02A1 viruses were selected for phylogenetic analysis with RAxML.

Antiretroviral (ARV) drug resistance in PR-RT sequences was analyzed with the online Stanford University HIV Drug Resistance Database's HIVdb Program[28].

To estimate the time of the most recent common ancestor (tMRCA) and to analyze the demographic growth of CRF63_02A1, we used a Bayesian Markov Chain Monte Carlo (MCMC) coalescent method as implemented in BEAST v1.7.4[29]. For these analyses, we used 105

CRF63_02A1 PR-RT sequences lacking drug resistance mutations, including 16 obtained by us and 89 downloaded from the HIV Sequence Database, collected between 2009 and 2013, as well as three CRF02_AG_{FSU} sequences used as outgroups. Since the evolutionary rate could not be inferred directly from CRF63_02A1 sequences, due to the narrow time span of sample collection, it was estimated using 100 A1 subsubtype PR-RT sequences (considering that most of the analyzed sequence is of A1 subsubtype) sampled along a time span of 23 years and lacking drug resistance associated mutations, retrieved from the HIV Sequence Database. In the subsequent analysis, the estimated posterior distribution of the substitution rate was incorporated as a prior distribution in the analysis of CRF63_02A1 sequences. The substitution model used for these analyses was HKY with gamma-distributed rate heterogeneity among sites and two partitions in the codon positions (1st+2nd, 3rd); other priors were an uncorrelated lognormal relaxed clock model and a Bayesian skyline plot demographic model. Each MCMC chain was run for 50 million generations, sampling every 1,000 generations, with the first 10% discarded as burn-in. MCMC convergence and effective samples sizes were checked using the program Tracer v.1.5.

At least one genome fragment could be sequenced in 24 HIV-1 samples from Novosibirsk: 16 in both PR-RT and the V3 region, 6 only in PR-RT, and 2 only in V3. Demographic and clinical data of these samples are shown in Table 2. Phylogenetic trees of PR-RT and V3 segments are shown in Fig. 1a and 1b, respectively. Phylogenetic classification according to these analyses was incorporated in Table 2. In PR-RT, where CRF63_02A1 viruses grouped in a monophyletic clade closely related to CRF02_AG_{FSU} viruses, 18 (81.8%) samples were CRF63_02A1 and 4 (18.2%) were A_{FSU} (Fig. 1a). In the V3 region, where CRF63_02A1 viruses did not group in a separate clade, but branched in a clade together with CRF02_AG_{FSU} viruses, 17 (94.4%) samples branched in the CRF02_AG_{FSU}/CRF63_02A1 clade and one (5.6%) was A_{FSU} (Fig. 1b). In the 16

samples in which both segments could be sequenced, 14 (87.5%) were CRF63_02A1 in both segments, one, RU_8508, was A_{FSU} in both segments, and one, RU_8506, had incongruent topologies, being A_{FSU} in PR-RT and CRF63_02A1 in V3, suggesting the presence of recombination between both variants. Of the 20 samples in which at least one segment was of CRF63_02A1 and data on transmission route was available, 13 (65%) corresponded to IDUs and 7 (35%) to heterosexual transmissions. The identification of CRF63_02A1 both among IDUs and heterosexually-infected individuals is similar to what has been reported with regard to A_{FSU} in Russia[8,9]. Epidemiological data indicate that heterosexual transmission of HIV-1 in FSU countries is closely linked to the epidemic among IDUs[1,30], which is consistent with our studies in St. Petersburg showing that infections among IDUs and heterosexually-infected individuals were not associated with different AFSU clusters[10]. With regard to CRF63_02A1, analyses of a greater number of samples with epidemiological data would be required to examine the existence of transmission networks and their possible associations with different transmission routes.

ARV drug resistance-associated mutations were found in three samples, RU_8148, RU_8499, and RU_8516. The first had low degree resistance to nucleoside RT inhibitors (M41L), the second had high or intermediate resistance to nonnucleoside RT inhibitors (K103N), and the third had resistance to both drug classes (D67N, K70R, M184V; K101E, G190S). RU_8148 and RU_8516 were on ARV drug treatment and no information on drug treatment was available for RU_8499.

Near full-length genome sequences were obtained in three samples. The phylogenetic tree (Fig. 1c) shows that the newly derived sequences branch within the CRF63_02A1 clade, which is closely related to the CRF02_AG_{FSU} clade. To analyze the mosaic structure of CRF63_02A1, a consensus sequence was obtained using all 11 available CRF63_02A1 near full-length genome sequences, including the 3 newly derived ones and the 8 obtained

previously[17,18], which was used for bootscan analysis. Since there is prior knowledge on the parental strains of CRF63_02A1[17] we used as references only consensus sequences of A_{FSU} and CRF02_AG_{FSU}, with consensuses of subtypes C and H being used as outgroups. The plot derived from this analysis (Fig. 2a) suggests a complex recombinant structure with most of the genome derived from CRF02_AG_{FSU}. Short segments (<200 nt) in which clustering with either A_{FSU} or CRF02_AG_{FSU} consensuses was supported by ≥50% bootstrap values were further analyzed via ML with RAxML and PhyML (Fig. 2b). These analyses allowed to assign 8 short segments to either variant clustering with the respective reference sequences with ≥70% bootstrap and/or ≥0.8 aLRT-SH-like support values. The mosaic structure of CRF63_02A1 derived from the bootscan analysis with Simplot combined with the ML trees of partial segments (Fig. 2c) shows the existence of 18 breakpoints delimiting 10 CRF02_AG_{FSU}-derived segments and 9 A_{FSU}-derived segments. The inferred structure is much more complex than that previously described[17], in which only five A_{FSU}-derived segments were identified. The difference may derive from the different methods of analysis and from the use as references in our study of the parental strains of CRF63_02A1 (A_{FSU} and CRF02_AG_{FSU}), which may allow for more precise definition of the recombinant structure than when using more distantly related references.

In order to identify CRF63_02A1 sequences deposited in the Los Alamos HIV Sequence Database, CRF63_02A1 PR-RT sequences were used for BLAST searches with subsequent phylogenetic analyses, as described in the methods section. This analysis revealed that in addition to being found in Novosibirsk, where they represented 59.9% of database sequences from this region collected in 2009-2013 (results not shown), CRF63_02A1 viruses were found in the Eastern area of Russia, near the border with China, in the cities of Khabarovsk (13 of 88 sequences, 14.8%) and Blagoveshchensk (2 of 40 sequences, 5%), and in the Ust-Kamenogorsk region in East Kazakhstan (3 of 14 sequences, 21.4%) (Fig. 3). This is in agreement with recent studies (published after submission of this manuscript) reporting the identification of

CRF63_02A1 in the Russian Far East[31] and Kazakhstan[32]. However, we could not find CRF63_02A1 among database viruses from Kyrgysztan, as suggested by other authors[17].

Through Bayesian MCMC analyses using PR-RT sequences obtained by us and deposited in databases, the tMRCA of CRF63_02A1 was estimated in 2006.8 [95% highest posterior density (HPD) interval 2005.8 – 2007.7]. Demographic growth estimated through the Bayesian Skyline Plot shows an exponential increase in the effective number of infections during a period comprising most of 2008 and the first half of 2009, with subsequent stabilization (Fig. 4). These results are consistent with the earliest known date of HIV diagnosis of a CRF63_02A1 infection in 2007 (Table 1) and of collection of the first sample harboring a CRF63_02A1 virus in 2009[27]. They are also in accord with the largest annual increase in new HIV-1 diagnoses recorded in the Novosibirsk region, which occurred in 2008 (Fig. 5).

In summary, this study shows that CRF63_02A1 is highly predominant in recently collected samples in Novosibirsk , where it is circulating among IDUs and via heterosexual contact and that its mosaic structure is much more complex than previously reported. CRF63_02A1 originated recently and expanded rapidly, with most of its expansion taking place along an approximate period of only one and a half years (Fig. 4). The rapid expansion of CRF63_02A1 is reminiscent of that reported for other HIV-1 genetic forms in established epidemics[33,34], supporting the need for continued molecular epidemiological surveillance of HIV-1, and its propagation to distant areas of Russia and to Kazakhstan indicates that CRF63_02A1 should be taken in consideration in the design of vaccines adapted to HIV-1 strains circulating in Russia and other FSU countries, whose HIV-1 epidemics are becoming increasingly complex by the introduction of new genetic forms, phylogenetic diversification within the established strains[10,35-37], and recombination between cocirculating variants, with generation of new CRFs[17].

Acknowledgments

We thank the personnel at the Genomic Unit of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain, for technical assistance in sequencing, and Bonnie Mathieson, from Office of AIDS Research, National Institutes of Health, USA, for her support of this study. This work was funded by Office of AIDS Research, National Institutes of Health, USA, through the training program “Molecular epidemiology of HIV-1 in Eastern Europe and its significance for vaccine development”.

Sequences were deposited in GenBank under accessions KJ197185-KJ197224.

References

1. UNAIDS: Report on the global AIDS epidemic 2010.Available at http://www.unaids.org/globalreport/global_report.htm.

2. Bobkov A, Cheingsong-Popov R, Garaev M, et al.: Identification of an env G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. AIDS 1994; 8:1649-1656.

3. Lukashov VV, Cornelissen MT, Goudsmit J, et al.: Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. AIDS 1995; 9:435-439.

4. Leitner T, Korovina G, Marquina S, Smolskaya T, Albert J: Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variants. AIDS Res Hum Retroviruses 1996; 12:1595-1603.

5. Thomson MM, Vázquez de Parga E, Vinogradova A, et al.: 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 2007; 23:1599-1604.

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

7. Nabatov AA, Kravchenko ON, Lyulchuk MG, Shcherbinskaya AM, Lukashov VV: Simultaneous introduction of HIV type 1 subtype A and B viruses into injecting drug users in southern

Ukraine at the beginning of the epidemic in the former Soviet Union. AIDS Res Hum Retroviruses 2002; 18:891-895.

8. Bobkov AF, Kazennova EV, Selimova LM, et al.: Temporal trends in the HIV-1 epidemic in Russia: predominance of subtype A. J Med Virol 2004; 74:191-196.

9. Vázquez de Parga E, Rakhmanova A, Pérez-Álvarez L, et al.: Analysis of drug resistance-associated mutations in treatment-naive individuals infected with different genetic forms of HIV-1 circulating in countries of the former Soviet Union. J Med Virol 2005; 77:337-344.

10. Thomson MM, Vinogradova A, Delgado E, et al.: Molecular epidemiology of HIV-1 in St Petersburg, Russia: predominance of subtype A, former Soviet Union variant, and identification of intrasubtype subclusters. J Acquir Immune Defic Syndr 2009; 51:332-339.

11. Liitsola K, Tashkinova I, Laukkanen T, et al.: HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. AIDS 1998; 12:1907-1919.

12. Smolskaya T, Liitsola K, Zetterberg V, et al.: HIV epidemiology in the Northwestern Federal District of Russia: dominance of HIV type 1 subtype A. AIDS Res Hum Retroviruses 2006; 22:1074-1080.

13. Fernández-García A, Cuevas MT, Vinogradova A, et al.: Near full-length genome characterization of a newly identified HIV type 1 subtype F variant circulating in St. Petersburg, Russia. AIDS Res Hum Retroviruses 2009; 25:1187-1191.

14. Carr JK, Nadai Y, Eyzaguirre L, et al.: Outbreak of a West African recombinant of HIV-1 in Tashkent, Uzbekistan. J Acquir Immune Defic Syndr 2005; 39:570-575.

15. Eyzaguirre LM, Erasilova IB, Nadai Y, et al.: Genetic characterization of HIV-1 strains circulating in Kazakhstan. J Acquir Immune Defic Syndr 2007; 46:19-23.

16. Laga VIu, Kazennova EV, Vasil'ev AV, et al.: [Molecular-genetic characterization of the HIV-1 variants abundant in Kirghizia]. Vopr Virusol 2012; 57:26-32.

17. Baryshev PB, Bogachev VV, Gashnikova NM: The HIV-1 genetic diversity in Russia: CRF63_02A1, a new HIV-1 genetic variant spreading in Siberia. AIDS Res Hum Retroviruses 2014; 30:592-597.

18. Baryshev PB, Bogachev VV, Gashnikova NM. Genetic characterization of an isolate of HIV type 1 AG recombinant form circulating in Siberia, Russia. Arch Virol 2012; 157:2335-2341.

19. Delwart EL, Shpaer EG, Louwagie J, et al.: Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. Science 1993; 262:1257-1261.

20. Delgado E, Thomson MM, Villahermosa ML, et al.: Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. J Acquir Immune Defic Syndr 2002; 29:536-543.

21. Sierra M, Thomson MM, Ríos M, et al.: The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. Infect Genet Evol 2005; 5:209-217.

22. Revilla A, Delgado E, Christian EC, et al.: Construction and phenotypic characterization of HIV type 1 functional envelope clones of subtypes G and F. AIDS Res Hum Retroviruses 2011; 27:889-901.

23. Katoh K, Standley DM: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013; 30:772-780.

24. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22:2688-2690.

25. Lole KS, Bollinger RC, Paranjape RS, et al.: Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 1999; 73:152-160.

26. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010; 59:307-321.

27. HIV Sequence Database. Los Alamos National Laboratory. Accessible at http://www.hiv-lanl.gov.

28. Shafer RW, Jung DR, Betts BJ. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. Nat Med 2000; 6:1290-1292.

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

30. Donoghoe MC, Matic S. HIV-1 in eastern Europe. Lancet 2003; 361: 1910-1911.

31. Kazennova E, Laga V, Lapovok I, et al.: HIV-1 genetic variants in the Russian Far East. AIDS Res Hum Retroviruses. 2014 May 15. [Epub ahead of print].

32. Lapovok I, Kazennova E, Laga V, et al.: Molecular epidemiology of HIV Type 1 infection in Kazakhstan: CRF02_AG prevalence is increasing in the Southeastern Provinces. AIDS Res Hum Retroviruses. 2014 Jun 19. [Epub ahead of print].

33. Pérez L, Thomson MM, Bleda MJ, et al.: HIV Type 1 molecular epidemiology in Cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. AIDS Res Hum Retroviruses 2006; 22:724-733.

34. Thomson MM, Fernández-García A, Delgado E, et al.: Rapid expansion of a HIV-1 subtype F cluster of recent origin among men who have sex with men in Galicia, Spain. J Acquir Immune Defic Syndr 2012; 59:e49-e51.

35. Roudinskii NI, Sukhanova AL, Kazennova EV, et al.: Diversity of human immunodeficiency virus type 1 subtype A and CRF03_AB protease in Eastern Europe: selection of the V77I variant and its rapid spread in injecting drug user populations. J Virol 2004; 78:11276-1187.

36. Fernández-García A, Revilla A, Vázquez de Parga E, et al.: The analysis of near full-length genome sequences of HIV type 1 subtype A viruses from Russia supports the monophyly of major intrasubtype clusters. AIDS Res Hum Retroviruses 2012; 28:1340-1343.

37. Vinogradova A, Gafurova E, Muñoz-Nieto M, Rakhmanova A, Osmanov S, Thomson MM. 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 2010; 26:395-400.

Figure legends

Fig. 1. Phylogenetic trees of HIV-1 sequences from samples from Novosibirsk. (a) Tree of PR-RT sequences. (b) Tree of *env* V3 region sequences. (c) Tree of CRF63_02A1 near full-length genome sequences. Samples from the current study are in bold type. Only bootstrap values ≥70% are shown. In references of subtype A, CRF02_AG and CRF63_02A1, country of sample collection is indicated with the ISO two-letter code.

Fig. 2. Analysis of the mosaic structure of CRF63_02A1. (a) Bootscan analysis of the consensus CRF63_02A1 near full-length genome sequence. Consensus A_{FSU} and CRF02_AG_{FSU} sequences were used as references, and subtype C and H as outgroups. The horizontal axis corresponds to positions in the HXB2 proviral genome sequence. (b) Phylogenetic trees of 8 short (<200 nt) putatively recombinant segments of CRF63_02A1, as identified in the Simplot bootscan analysis [signaled in (a)]; nucleotide positions in the HXB2 genome are indicated on top of each tree in parentheses; trees are rooted with the inferred group M ancestral sequence; branches corresponding to CRF63_02A1 viruses are marked with filled circles, those of CRF02_AG_{FSU} viruses with triangles, and those of A_{FSU} viruses with squares; bootstrap support, as determined with RAxML, and aLRT-SH-like support, as determined with PhyML, of A_{FSU} and CRF02_AG_{FSU} clades are shown on the left of the corresponding nodes, above and below, respectively, of the subtending branch. (c) Mosaic structure of CRF63_02AG inferred from Simplot analysis and ML trees of partial segments. HXB2 positions of breakpoints delimiting CRF02_AG_{FSU}-derived and A_{FSU}-derived segments are indicated on top.

Fig. 3. Phylogenetic tree of PR-RT sequences showing viruses from East Kazakhstan and the Eastern Russian cities of Khabarovsk and Blagoveshchensk branching within the CRF63_02A1

clade. The names of the mentioned viruses, which include the KAZ, KHA and BLG abbreviations of the sampling locations, are in bold type. Only bootstrap values ≥70% are shown.

Fig. 4. Bayesian skyline plot of the population growth of CRF63_02A1. The plot was obtained using PR-RT sequences. The black line represents the median estimate of the effective number of infections through time, plotted on a logarithmic scale, and the shaded area represents the 95% HPD confidence interval for this estimate.

Fig. 5. Time trend in new HIV-1 diagnoses in the Novosibirsk region (2002-2012). Data are from the Center for Prevention and Control of AIDS and Infectious Diseases, Koltsovo, Novosibirsk region, Russia (http://spidmso.ru).

Fig. 1

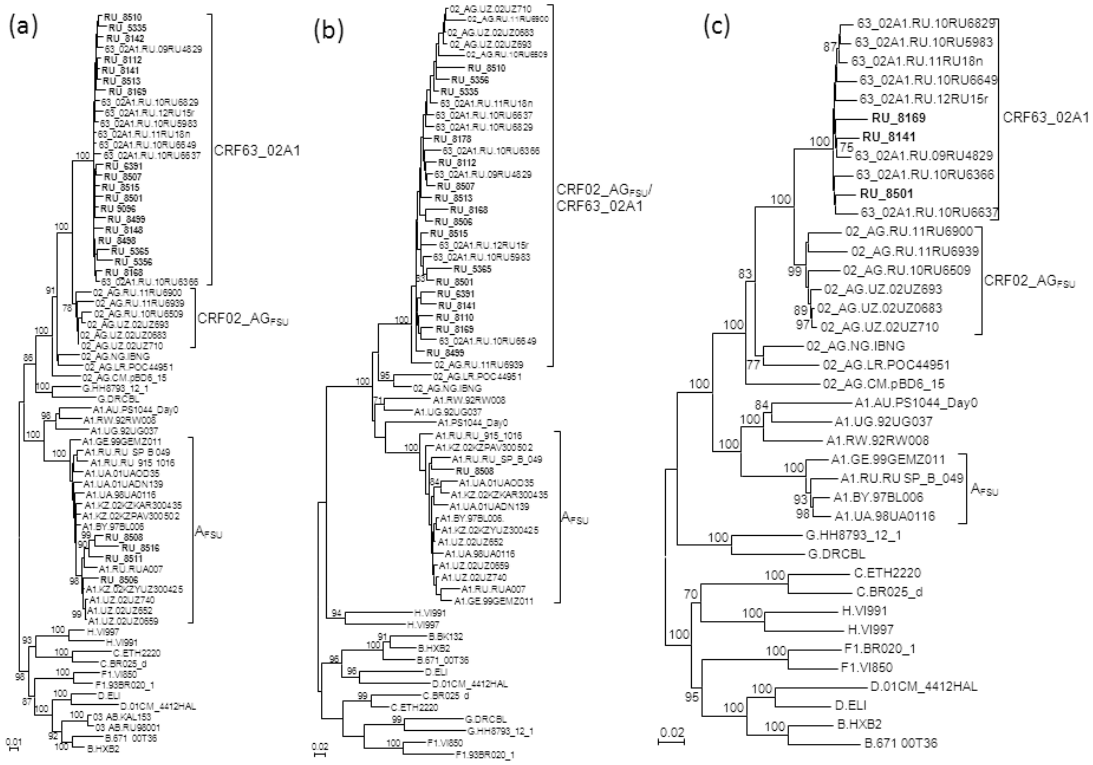

(a)
CRF63_02A1
CRF02_AG$_{FSU}$
A$_{FSU}$
(b)
CRF02_AG$_{FSU}$/
CRF63_02A1
A$_{FSU}$
(c)
CRF63_02A1
CRF02_AG$_{FSU}$
A$_{FSU}$

Fig. 2

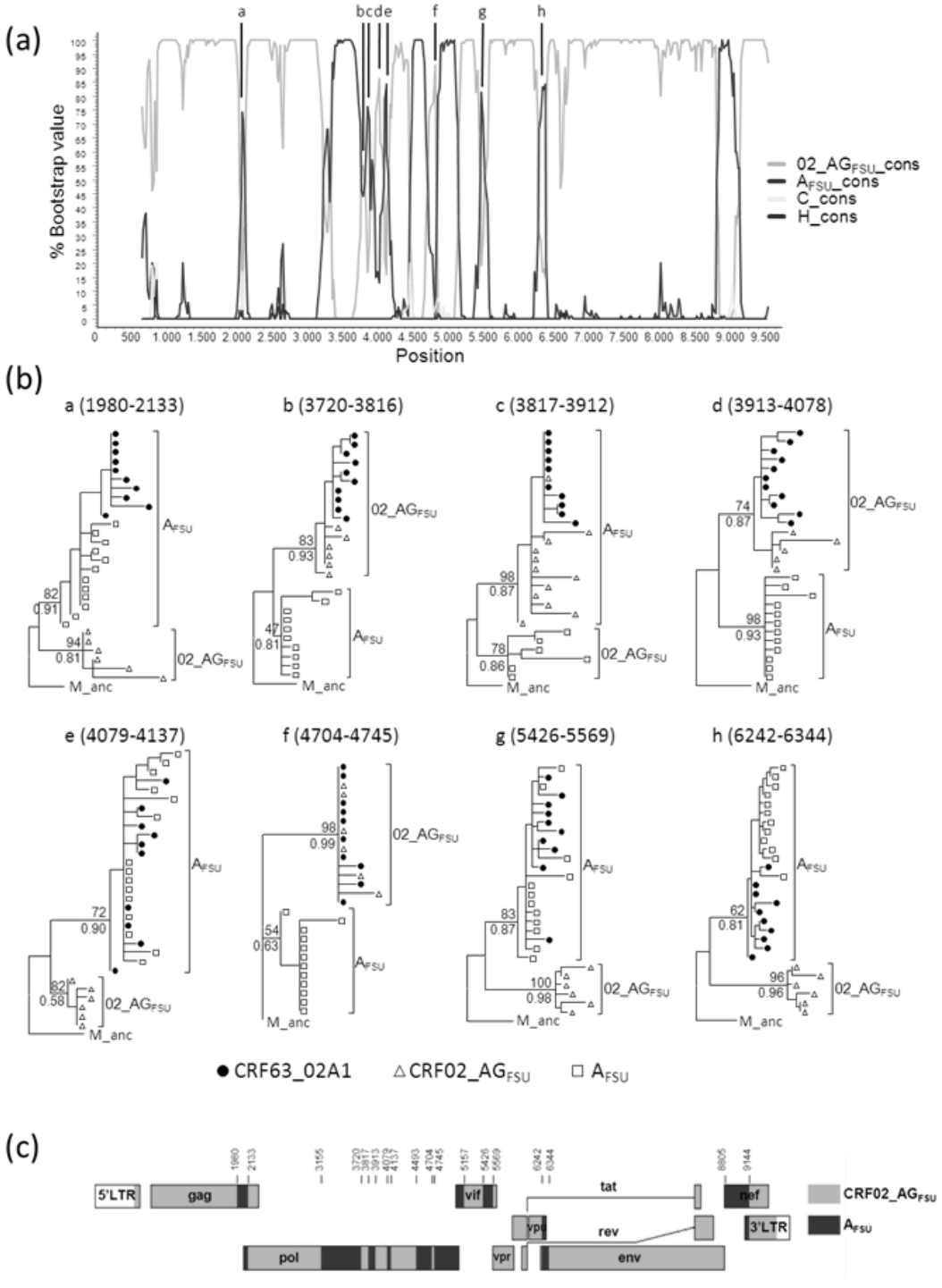

(a)
a
bcde
f
g
h
% Bootstrap value
Position
02_AG_FSU_cons
A_FSU_cons
C_cons
H_cons
(b)
a (1980-2133)
b (3720-3816)
c (3817-3912)
d (3913-4078)
e (4079-4137)
f (4704-4745)
g (5426-5569)
h (6242-6344)
A_FSU
02_AG_FSU
M_anc
● CRF63_02A1
△ CRF02_AG_FSU
□ A_FSU
(c)
5'LTR
gag
pol
vif
vpr
tat
rev
env
nef
3'LTR
CRF02_AG_FSU
A_FSU

Fig. 3

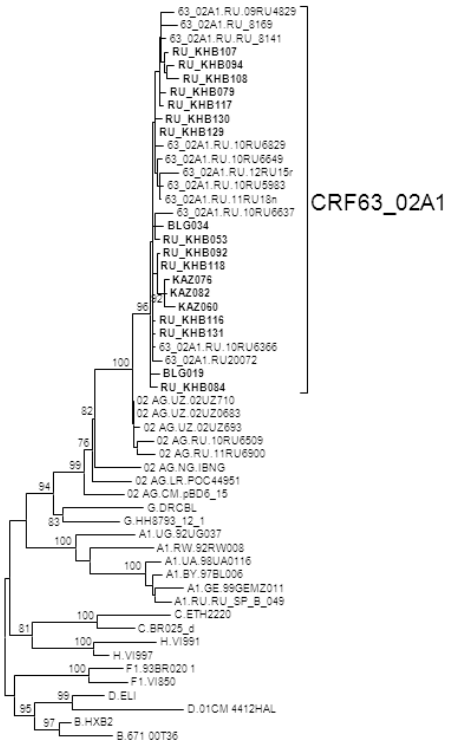

63_02A1.RU.09RU4829
63_02A1.RU_8169
63_02A1.RU.RU_8141
RU_KHB107
RU_KHB094
RU_KHB108
RU_KHB079
RU_KHB117
RU_KHB130
RU_KHB129
63_02A1.RU.10RU6829
63_02A1.RU.10RU6649
63_02A1.RU.12RU15r
63_02A1.RU.10RU5983
63_02A1.RU.11RU18n
63_02A1.RU.10RU6637
BLG034
RU_KHB053
RU_KHB092
RU_KHB118
KAZ076
KAZ082
KAZ060
RU_KHB116
RU_KHB131
63_02A1.RU.10RU6366
63_02A1.RU20072
BLG019
RU_KHB084
CRF63_02A1
02.AG.UZ.02UZ710
02.AG.UZ.02UZ0683
02.AG.UZ.02UZ693
02.AG.RU.10RU6509
02.AG.RU.11RU6900
02.AG.NG.IBNG
02.AG.LR.POC44951
02.AG.CM.pBD6_15
G.DRCBL
G.HH8793_12_1
A1.UG.92UG037
A1.RW.92RW008
A1.UA.98UA0116
A1.BY.97BL006
A1.GE.99GEMZ011
A1.RU.RU_SP_B_049
C.ETH2220
C.BR025_d
H.VI991
H.VI997
F1.93BR020 1
F1.VI850
D.ELI
D.01CM 4412HAL
B.HXB2
B.671 00T36
96
100
82
76
99
94
83
100
100
100
81
100
100
99
95
97

Fig. 4

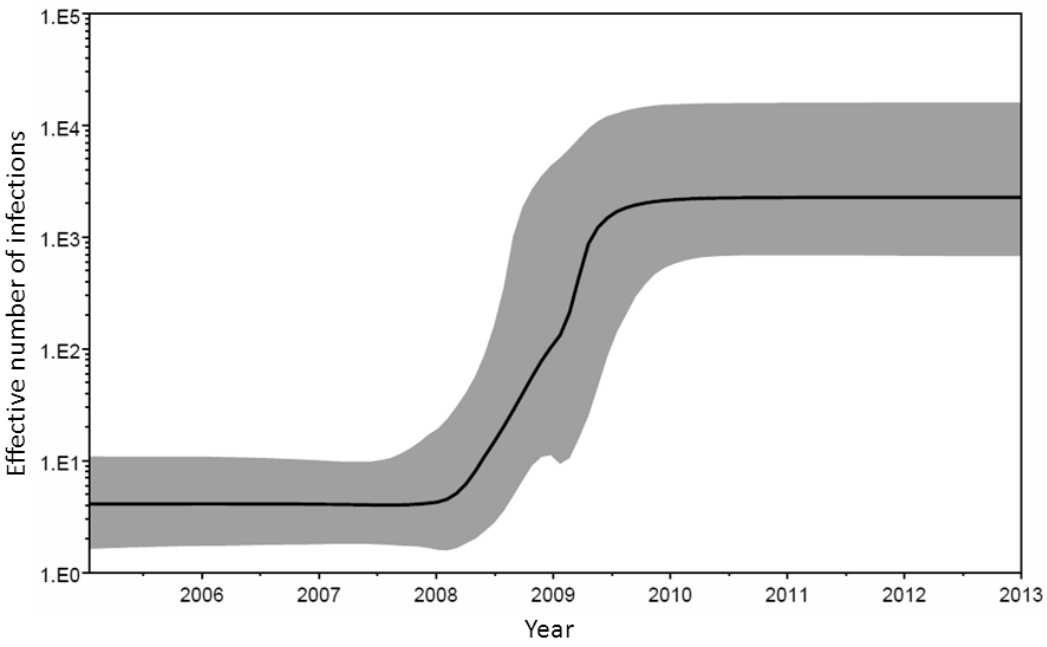
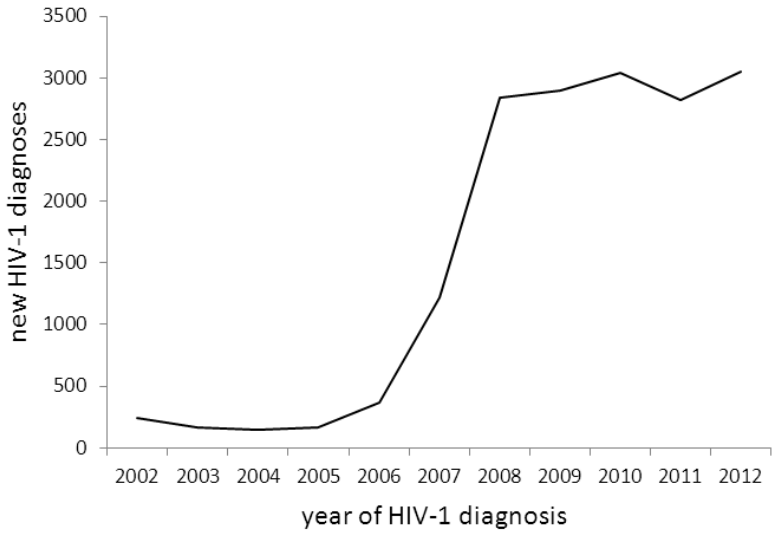

1.E5
1.E4
1.E3
1.E2
1.E1
1.E0
Effective number of infections
2006
2007
2008
2009
2010
2011
2012
2013
Year

Fig. 5


3500
3000
2500
2000
1500
1000
500
0
new HIV-1 diagnoses
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
year of HIV-1 diagnosis