

## Short Communication: Molecular Epidemiology, Phylogeny, and Phylodynamics of CRF63\_02A1, a Recently Originated HIV-1 Circulating Recombinant Form Spreading in Siberia

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

### 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 three 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.

**R**USSIA IS ONE OF THE NON-AFRICAN COUNTRIES with the greatest number of HIV-1-infected persons, only surpassed by India and the United States.<sup>1</sup> Before 1996 there were approximately 1,000 HIV-1-infected people in Russia ([www.hivrussia.org/](http://www.hivrussia.org/)). In that period, subtype B was predominant, although other HIV-1 clades, of diverse geographic origins, were also detected.<sup>2–4</sup>

In 1995–1996 HIV-1 infections began to increase dramatically concomitantly with the expansion of a subtype A variant of Central African ancestry<sup>5</sup> that originated in Southern Ukraine<sup>6,7</sup> and spread to all countries of the former Soviet Union (FSU), in most of which it is the predominant HIV-1 genetic form<sup>8,9</sup>; for this reason it is frequently designated  $A_{FSU}$  variant. In addition to  $A_{FSU}$ , other HIV-1 genetic forms cir-

culating in FSU countries at lower prevalences include subtype B, predominant in men who have sex with men,<sup>8–10</sup> CRF03\_AB, predominant in the Russian cities of Kaliningrad<sup>11</sup> and Cherepovets,<sup>12</sup> subtype F, circulating as a minor variant in St. Petersburg, Russia,<sup>13</sup> and a CRF02\_AG variant (CRF02\_AG<sub>FSU</sub>),<sup>14,15</sup> which was first detected among injecting drug users (IDUs) in Tashkent, Uzbekistan, in 1999–2000,<sup>14</sup> and has subsequently been reported in Kazakhstan,<sup>15</sup> Kyrgyzstan,<sup>16</sup> and in the Novosibirsk region, Siberia, Russia.<sup>17</sup> This variant has generated, through secondary recombination with  $A_{FSU}$ , a new circulating recombinant form, CRF63\_02A1, recently identified in Novosibirsk.<sup>17,18</sup> Here we examine the prevalence of CRF63\_02A1 in the Novosibirsk region in samples collected in 2013, we obtain three

<sup>1</sup>Department of Bioengineering, State Research Center of Virology and Biotechnology “Vector,” Koltsovo, Novosibirsk, Russia.

<sup>2</sup>Department of Zoonotic Infections and Influenza, State Research Center of Virology and Biotechnology “Vector,” Koltsovo, Novosibirsk, Russia.

<sup>3</sup>HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain.

<sup>4</sup>Laboratory of Recombinant Vaccine, State Research Center of Virology and Biotechnology “Vector,” Koltsovo, Novosibirsk, Russia.

<sup>5</sup>Center for Prevention and Control of AIDS and Infectious Diseases, Koltsovo, Novosibirsk, Russia.

<sup>6</sup>Altai State University, Barnaul, Russia.

\*These authors contributed equally to this work.

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 reverse transcriptase polymerase chain reaction (RT-PCR) followed by nested PCR. Primers used for PR-RT amplification were RP1-S and RP1-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.<sup>19</sup> 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,<sup>20,21</sup> 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.<sup>22</sup>

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; 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).

Sequences were aligned with MAFFT v.7.<sup>23</sup> Phylogenetic trees were constructed via maximum likelihood with RAxML v.7.2.7,<sup>24</sup> 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,<sup>25</sup> 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 ( $\leq 200$  nt) identified with SimPlot were phylogenetically analyzed further via maximum likelihood (ML) using RAxML, applying the GTR + CAT model, and with PhyML v3.0,<sup>26</sup> 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<sub>FSU</sub> and A<sub>FSU</sub> and were rooted with an inferred group M ancestral sequence downloaded from the HIV Sequence Database.<sup>27</sup> Segments were assigned to either variant based on clustering with a bootstrap value  $\geq 70\%$  with RAxML or an aLRT-SH-like support  $\geq 0.8$  with PhyML.

To identify CRF63\_02A1 viruses from other geographic areas, we downloaded all HIV-1 sequences from FSU countries

TABLE 1. PRIMERS USED FOR POLYMERASE CHAIN REACTION AMPLIFICATION OF HIV-1 NEAR FULL-LENGTH GENOMES

| Name              | Sequence                          | HXB2 positions | Orientation <sup>a</sup> |
|-------------------|-----------------------------------|----------------|--------------------------|
| SC-A-O-S-R        | AGATCTGAGCCTGGGAGCTCTCTGG         | 473–497        | S                        |
| SC-A-O-A          | TTTATGGCAAATACTGGAGTATTGTATGGA    | 2711–2740      | A                        |
| SC-A-N-S-R        | GCTTCAAGTAGTGTGTGCCCGTCTG         | 546–570        | S                        |
| SC-A-N-A          | CTTTTGGGCCATCCATTCCTG             | 2590–2610      | A                        |
| RP1-S             | GAAAAAGGGCTGTTGGAAATGTGGAA        | 2016–2041      | S                        |
| RP1-A             | AAATTTAGGAGTCTT TCCCATATTACTATGC  | 3685–3716      | A                        |
| PR-O-S2b          | GCTAATTTT TTAGGGAARATYTGCCCT T    | 2080–2107      | S                        |
| RT-O-A            | TGCCCTGTTAATTGTTTTACATCATTAGTGTG  | 3630–3662      | A                        |
| RTDS-O-S          | ATCTATCAATACATGGATGATTGTATGTAG    | 3087–3117      | S                        |
| SC-B-N-A          | CCCTAGTGGGATGTGTACTTCTGAACTTA     | 5192–5220      | A                        |
| RTDS-N-S          | ATAGGGCAGCATAGARCAAAAATAGA        | 3132–3157      | S                        |
| B2-N-A            | AATCATCACCTGCCATCTGTTTTCC         | 5043–5067      | A                        |
| SG3-up            | TACAGTGCAGGGGAAAGAATAATAGACATAATA | 4809–4841      | S                        |
| SG3-lo            | AGACCCAGTACAGGCRARAAGC            | 9523–9544      | A                        |
| Env-S             | AGGCATCTCCTATGGCAGGAAGAAG         | 5957–5983      | S                        |
| Env-A             | CCAGTCCCCCTTTTCTTTTAAAAAAG        | 9063–9088      | A                        |
| SG3-N-S           | CAAAATTTTCGGGTTTATTACAGGGACAG     | 4890–4918      | S                        |
| SSD2b             | TCTYTCCACACAGGTACCCATA            | 6342–6364      | A                        |
| NEF-S4            | AGTATTGGAGTCAGGAATAAAGAATAGTG     | 8626–8655      | S                        |
| 3' <i>nef</i> -3c | GTACAGGCRARAAGCAGCTGCTTATATG      | 9510–9537      | A                        |

<sup>a</sup>S, sense; A, antisense.

deposited in the HIV Sequence Database.<sup>27</sup> The 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.<sup>28</sup>

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.<sup>29</sup> 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<sub>FSU</sub> 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 subtype PR-RT sequences (considering that most of the analyzed sequence is of A1 subtype) 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 het-

erogeneity among sites and two partitions in the codon positions (first+second, third); 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 sample 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, six only in PR-RT, and two 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 b, 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<sub>FSU</sub> viruses, 18 (81.8%) samples were CRF63\_02A1 and four (18.2%) were A<sub>FSU</sub> (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<sub>FSU</sub> viruses, 17 (94.4%) samples branched in the CRF02\_AG<sub>FSU</sub>/CRF63\_02A1 clade and one (5.6%) was A<sub>FSU</sub> (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<sub>FSU</sub> in both segments, and one, RU\_8506, had incongruent topologies, being A<sub>FSU</sub> in PR-RT and CRF63\_02A1 in V3, suggesting the presence of recombination between both variants.

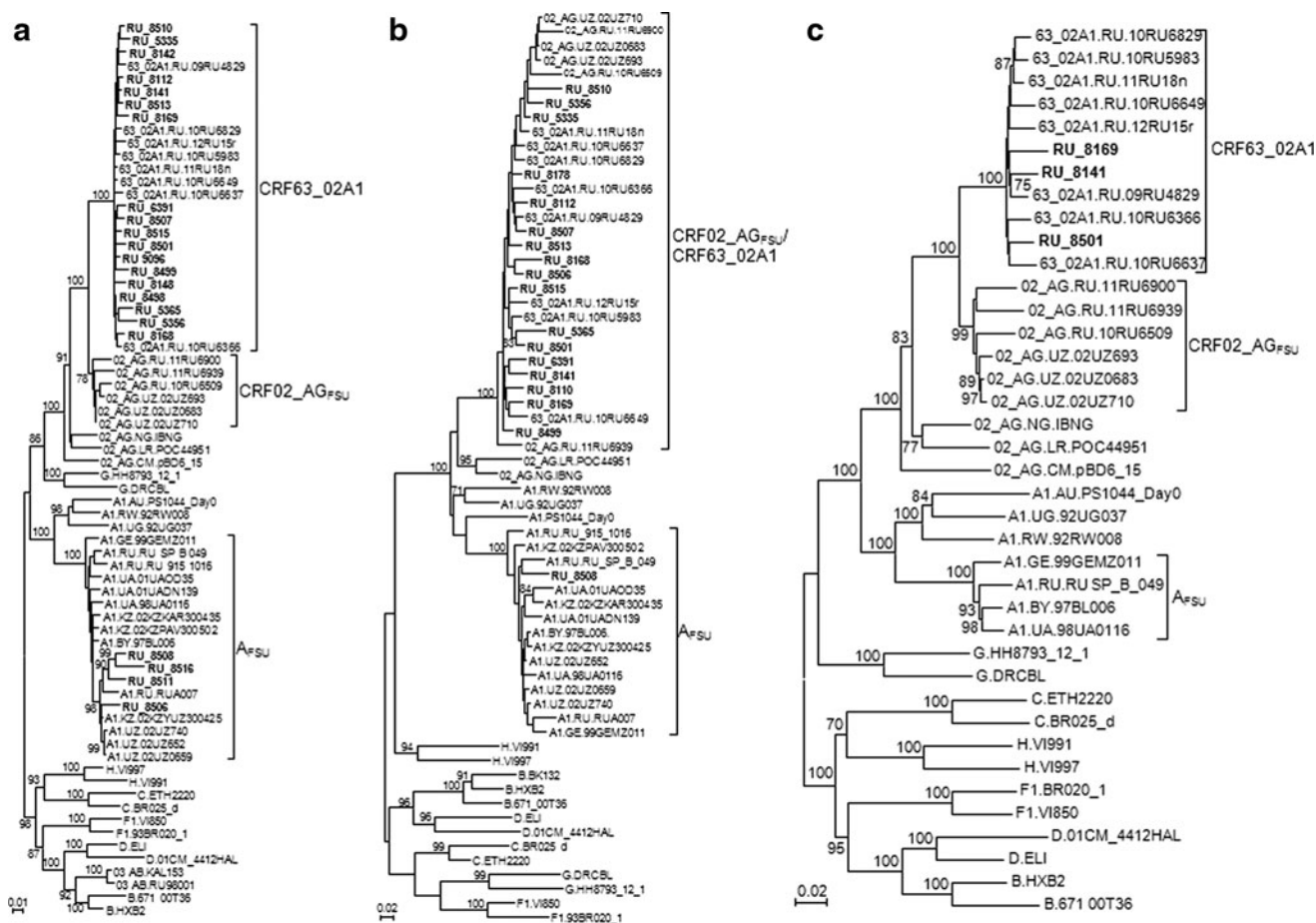
Of the 20 samples in which at least one segment was of CRF63\_02A1 and data on transmission route were available,

TABLE 2. DATA OF SAMPLES WHOSE SEQUENCES WERE ANALYZED IN THIS STUDY, WITH SUBTYPE/CIRCULATING RECOMBINANT FORM CLASSIFICATION IN PROTEASE REVERSE TRANSCRIPTASE AND V3 FRAGMENTS

| Sample ID | Gender | Transmission route <sup>a</sup> | Year of HIV diagnosis | ARV treatment | Subtype/CRF PR-RT | Subtype/CRF V3 |
|-----------|--------|---------------------------------|-----------------------|---------------|-------------------|----------------|
| RU_5335   | M      | Het                             | 2013                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_5356   | n.a.   | n.a.                            | n.a.                  | n.a.          | 63_02A1           | 63_02A1/02_AG  |
| RU_5365   | F      | Het                             | 2013                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_6391   | F      | Het                             | 2013                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8110   | F      | Het                             | 2011                  | No            |                   | 63_02A1/02_AG  |
| RU_8112   | M      | Het                             | 2012                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8141   | M      | IDU                             | 2007                  | Yes           | 63_02A1           | 63_02A1/02_AG  |
| RU_8142   | F      | IDU                             | 2011                  | Yes           | 63_02A1           |                |
| RU_8148   | M      | IDU                             | 2011                  | Yes           | 63_02A1           |                |
| RU_8168   | M      | IDU                             | 2010                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8169   | M      | IDU                             | 2009                  | n.a.          | 63_02A1           | 63_02A1/02_AG  |
| RU_8178   | M      | IDU                             | 2009                  | Yes           |                   | 63_02A1/02_AG  |
| RU_8498   | M      | IDU                             | 2009                  | No            | 63_02A1           |                |
| RU_8499   | M      | IDU                             | n.a.                  | n.a.          | 63_02A1           | 63_02A1/02_AG  |
| RU_8501   | F      | Het                             | 2012                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8506   | M      | IDU                             | 2012                  | No            | A1                | 63_02A1/02_AG  |
| RU_8507   | M      | IDU                             | 2012                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8508   | F      | Het                             | 2013                  | No            | A1                | A1             |
| RU_8510   | F      | IDU                             | 2011                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8511   | M      | Sex                             | 2012                  | No            | A1                |                |
| RU_8513   | F      | IDU                             | 2012                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8515   | F      | IDU                             | 2011                  | No            | 63_02A1           | 63_02A1/02_AG  |
| RU_8516   | M      | IDU                             | 2012                  | Yes           | A1                |                |
| RU_9096   | F      | Het                             | 2006                  | No            | 63_02A1           |                |

<sup>a</sup>IDU, injecting drug user; Het, heterosexual; Sex, sexual.

n.a., datum not available; ARV, antiretroviral; CRF, circulating recombinant form; PR-RT, protease reverse transcriptase.



**FIG. 1.** Phylogenetic trees of HIV-1 sequences from samples from Novosibirsk. **(a)** Tree of the protease-reverse transcriptase (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  $\geq 70\%$  are shown. In references of subtype A, CRF02\_AG, and CRF63\_02A1, the country of sample collection is indicated by the ISO two-letter code.

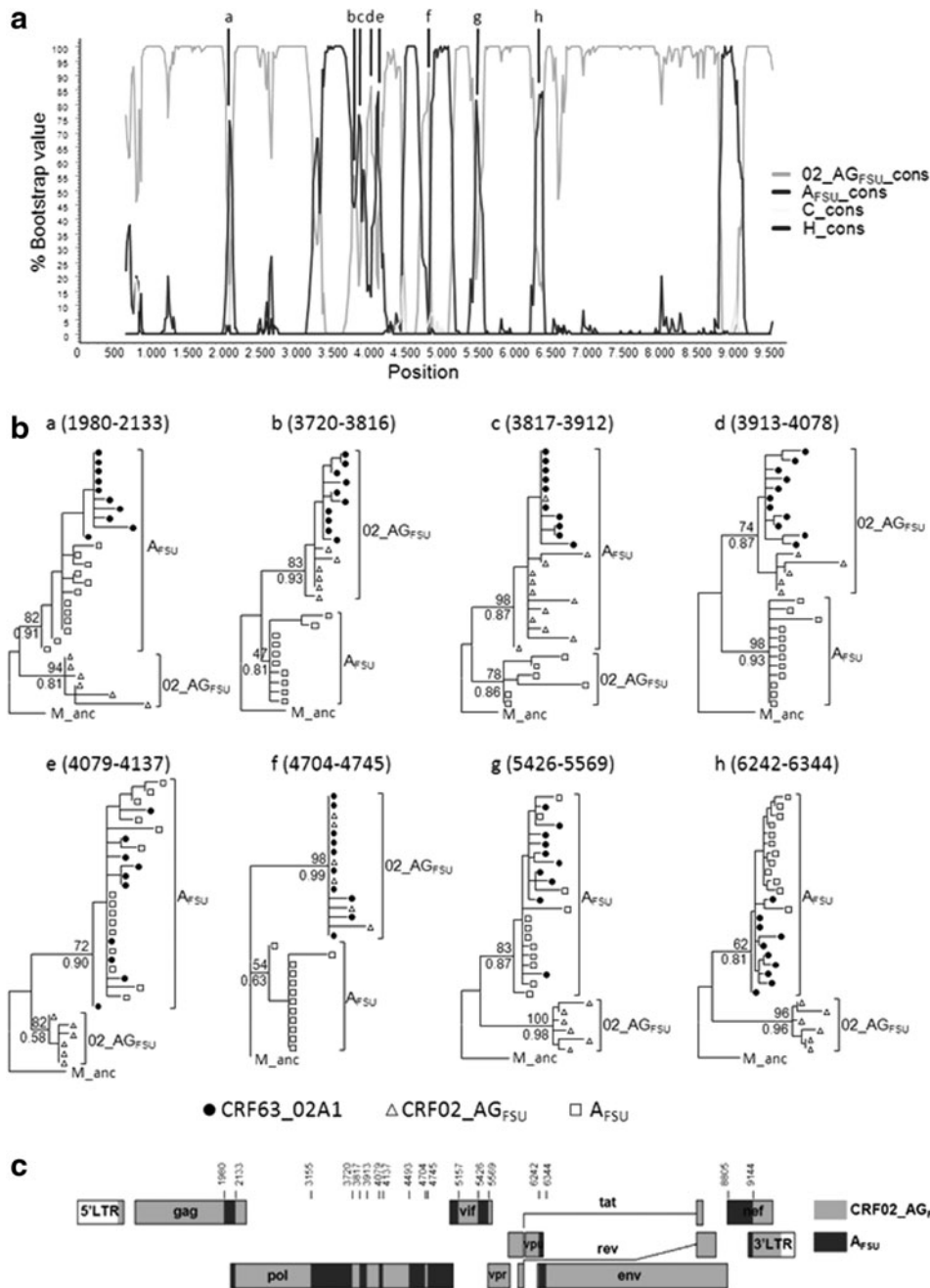
13 (65%) corresponded to IDUs and seven (35%) to heterosexual transmissions. The identification of CRF63\_02A1 both among IDUs and heterosexually infected individuals is similar to what has been reported with regard to A<sub>FSU</sub> in Russia.<sup>8,9</sup> Epidemiological data indicate that heterosexual transmission of HIV-1 in FSU countries is closely linked to the epidemic among IDUs,<sup>1,30</sup> which is consistent with our studies in St. Petersburg showing that infections among IDUs and heterosexually infected individuals were not associated with different A<sub>FSU</sub> clusters.<sup>10</sup> 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.

Antiretroviral (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<sub>FSU</sub> 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 three newly derived ones and the eight obtained previously,<sup>17,18</sup> which was used for bootscan analysis. Since there is prior knowledge on the parental strains of CRF63\_02A1<sup>17</sup> we used as references only consensus sequences of A<sub>FSU</sub> and CRF02\_AG<sub>FSU</sub>, with consensus 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<sub>FSU</sub>. Short segments (< 200 nt) in which clustering with either A<sub>FSU</sub> or CRF02\_AG<sub>FSU</sub> consensus was supported by  $\geq 50\%$  bootstrap values were further analyzed via ML with RAXML and PhyML (Fig. 2b). These analyses made it possible to assign eight short segments to either variant clustering with the respective reference sequences with  $\geq 70\%$  bootstrap and/or  $\geq 0.8$  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<sub>FSU</sub>-derived segments and nine A<sub>FSU</sub>-derived segments. The inferred structure is



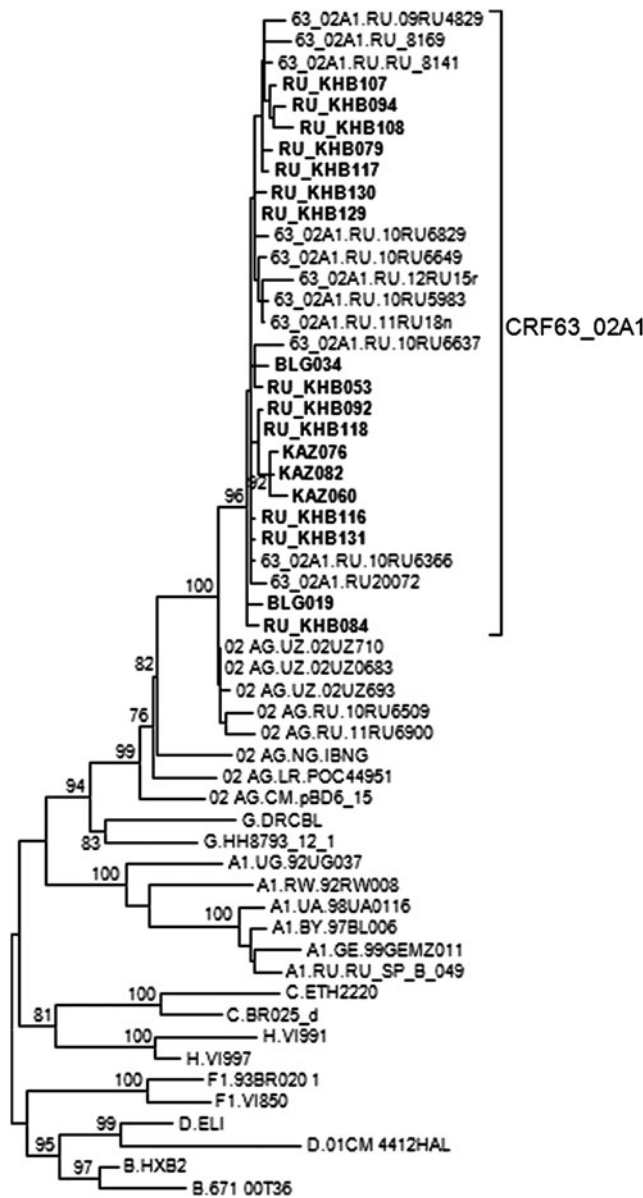
**FIG. 2.** Analysis of the mosaic structure of CRF63\_02A1. **(a)** Bootscan analysis of the consensus CRF63\_02A1 near full-length genome sequence. Consensus A<sub>F<sub>SU</sub></sub> and CRF02\_AG<sub>F<sub>SU</sub></sub> sequences were used as references and subtypes C and H as out-groups. The horizontal axis corresponds to positions in the HXB2 proviral genome sequence. **(b)** Phylogenetic trees of eight 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<sub>F<sub>SU</sub></sub> viruses with triangles, and those of A<sub>F<sub>SU</sub></sub> viruses with squares; bootstrap support, as determined with RAxML, and aLRT-SH-like support, as determined with PhyML, of A<sub>F<sub>SU</sub></sub> and CRF02\_AG<sub>F<sub>SU</sub></sub> 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. The HXB2 positions of breakpoints delimiting CRF02\_AG<sub>F<sub>SU</sub></sub>-derived and A<sub>F<sub>SU</sub></sub>-derived segments are indicated on the top.

much more complex than what was previously described,<sup>17</sup> in which only five A<sub>F<sub>SU</sub></sub>-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<sub>F<sub>SU</sub></sub> and CRF02\_AG<sub>F<sub>SU</sub></sub>), which may allow for a more precise definition of the recombinant structure than when using more distantly related references.

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 Materials and 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 reporting the identification of CRF63\_02A1 in the Russian Far East<sup>31</sup> and Kazakhstan.<sup>32</sup> However, we could not find CRF63\_02A1 among database viruses from Kyrgyzstan, as suggested by others.<sup>17</sup>

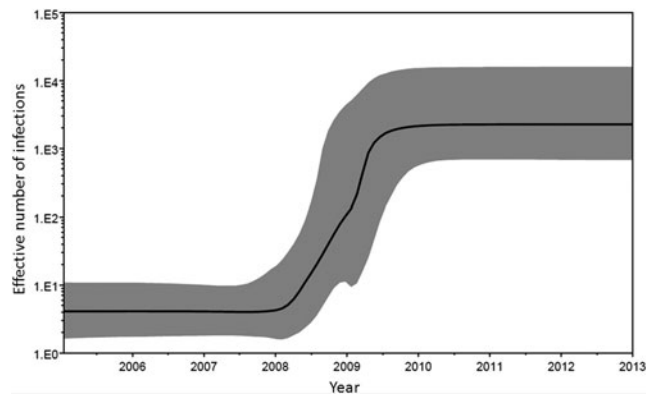
Through Bayesian MCMC analyses using PR-RT sequences obtained by us and deposited in databases, the tMRCA of CRF63\_02A1 was estimated in 2006.8 [95% highest posterior density (HPD) interval 2005.8–2007.7]. Demographic growth estimated through the Bayesian Skyline Plot shows an exponential increase in the effective



**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  $\geq 70\%$  are shown.

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 the collection of the first sample harboring a CRF63\_02A1 virus in 2009.<sup>27</sup> 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).

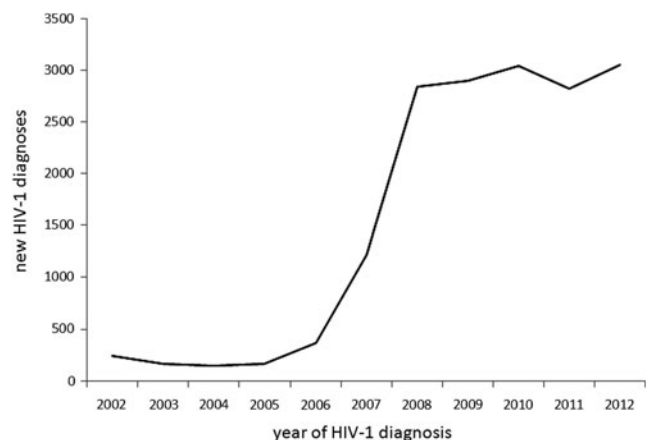
In summary, this study shows that CRF63\_02A1 is highly predominant in recently collected samples in Novosibirsk, where it is transmitted both among IDUs and via heterosexual contact and that its mosaic structure is much more complex



**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.

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,<sup>33,34</sup> supporting the need for continued molecular epidemiological surveillance of HIV-1. Its propagation to distant areas of Russia and to Kazakhstan indicates that CRF63\_02A1 should be taken into 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,<sup>10,35–37</sup> and recombination between cocirculating variants, with the generation of new circulating recombinant forms (CRFs).<sup>17</sup>

Sequences were deposited in GenBank under accession numbers KJ197185–KJ197224.



**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>).

### 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 the Office of AIDS Research, National Institutes of Health, Bethesda, Maryland for her support of this study. This work was funded by Office of AIDS Research, National Institutes of Health, through the training program “Molecular Epidemiology of HIV-1 in Eastern Europe and Its Significance for Vaccine Development.”

### Author Disclosure Statement

No competing financial interests exist.

### References

- UNAIDS: Report on the global AIDS epidemic 2010. Available at [www.unaids.org/globalreport/global\\_report.htm](http://www.unaids.org/globalreport/global_report.htm).
- 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.
- 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.
- Leitner T, Korovina G, Marquina S, Smolskaya T, and Albert J: Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variants. *AIDS Res Hum Retroviruses* 1996;12:1595–1603.
- 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.
- Novitsky VA, Montano MA, and 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.
- Nabatov AA, Kravchenko ON, Lyulchuk MG, Shcherbinskaya AM, and 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Laga Vlu, Kazennova EV, Vasil’ev AV, *et al.*: [Molecular-genetic characterization of the HIV-1 variants abundant in Kirghizia]. *Vopr Virusol* 2012;57:26–32.
- Baryshev PB, Bogachev VV, and 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.
- Baryshev PB, Bogachev VV, and Gashnikova NM: Genetic characterization of an isolate of HIV type 1 AG recombinant form circulating in Siberia, Russia. *Arch Virol* 2012;157:2335–2341.
- 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.
- 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.
- 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.
- 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.
- Katoh K and Standley DM: MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
- Stamatakis A: RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;22:2688–2690.
- 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.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, and Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
- HIV Sequence Database: Los Alamos National Laboratory. Accessible at [www.hiv-lanl.gov](http://www.hiv-lanl.gov).
- Shafer RW, Jung DR, and 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, and Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;29:1969–1973.
30. Donoghoe MC and Matic S: HIV-1 in eastern Europe. *Lancet* 2003;361:1910–1911.
31. Kazenova E, Laga V, Lapovok I, *et al.*: HIV-1 genetic variants in the Russian Far East. *AIDS Res Hum Retroviruses*. 2014;30:742–752.
32. Lapovok I, Kazenova 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;30:769–774.
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 inter-subtype recombinant forms. *AIDS Res Hum Retroviruses* 2006;22:724–733.
34. Thomson MM, Fernández-García A, Delgado E, *et al.*: Rapid expansion of an 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, Kazenova 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, and 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.

Address correspondence to:

*Michael M. Thomson*

*Centro Nacional de Microbiología*

*Instituto de Salud Carlos III*

*Carretera Majadahonda-Pozuelo, Km. 2*

*28220 Majadahonda Madrid*

*Spain*

*E-mail: mthomson@isciii.es*