The Analysis of Near Full-Length Genome Sequences of HIV Type 1 Subtype A Viruses from Russia Supports the Monophyly of Major Intrasubtype Clusters

Aurora Fernández-García,^{1,*} Ana Revilla,^{1,*} Elena Vázquez-de Parga,¹ Anna Vinogradova,² Aza Rakhmanova,³ Eduard Karamov,⁴ Cristina Carrera,¹ Elena Delgado,¹ Lucía Pérez-Álvarez,¹ Rafael Nájera,¹ Saladin Osmanov,⁵ and Michael M. Thomson¹

Abstract

The HIV-1 epidemic in Russia has been insufficiently studied, with only 11 complete genome sequences from this country currently available, only three of which are of the locally predominant genetic form, the former Soviet Union (FSU) subtype A variant (A_{FSU}). Here we analyze 10 newly derived A_{FSU} near full-length genome sequences from Russia. Samples were selected based on phylogenetic clustering in protease-reverse transcriptase in two of the major A_{FSU} clusters, $V77I_{PR}$ ($n=6$), widely circulating in Russia and other FSU countries, and A_{SP1} ($n=4$), predominant in St. Petersburg. The phylogenetic analysis shows that the $V77I_{PR}$ genomes group in a monophyletic cluster together with 10 previously obtained A_{FSU} genome sequences from Uzbekistan, Kazakhstan, Russia, and Cyprus, all bearing the $V77I$ substitution in protease. Similarly, the four A_{SP1} genomes group in a monophyletic cluster. These results therefore show that the monophyly of $V77I_{PR}$ and A_{SP1} A_{FSU} clusters is supported in near complete genomes.

THE HIV-1 EPIDEMIC IN EASTERN European and Central Asian countries formerly part of the Soviet Union is dominated by a subtype A monophyletic variant^{1,2} of Central African origin that, after an initial limited propagation via heterosexual contact,^{3,4} began spreading explosively among injecting drug users (IDU) in the Ukrainian seaport city of Odessa in early 1995.⁵ This variant is the predominant HIV-1 genetic form in all former Soviet Union (FSU) countries studied except in Estonia, and for this reason it is often referred to as the A_{FSU} variant. The number of people infected with HIV-1 in Eastern European and Central Asian countries where the A_{FSU} variant is circulating was estimated to be approximately 1.4 million at the end of 2009.⁶

The greatest burden of the epidemic in this region is in the Russian Federation with 980,000 estimated cases at the end of 2009 (the third largest number of HIV-1-infected people among non-African countries after India and the United States) and a prevalence of 1%, followed by Ukraine, with 350,000 estimated cases and a 1.1% prevalence. In spite of its magnitude, the HIV-1 epidemic in FSU countries has been

insufficiently studied from a genetic point of view. Thus, at the Los Alamos HIV Sequence database, only 3948 HIV-1 sequences from FSU countries are available, which is very low compared to those from Western and Central Europe, with 81,757 sequences for an estimated number of 820,000 infections, North America, with 126,795 sequences for 1.5 million estimated infections, and South and Central America, with 21,957 sequences for 1.4 million estimated infections.⁷ The number of HIV-1 full-length or near full-length genome sequences available from FSU countries follows a similar trend, with only 47 sequences available from these countries, compared to 438 from Western and Central Europe, 622 from North America, and 242 from South and Central America. Considering HIV-1 isolates from Russia, only 11 near full-length genome sequences are available, only three of which are from the A_{FSU} variant, which represents more than 90% HIV-1 infections in the country.²

The scarcity of available near full-length HIV-1 genome sequences from Russia was one of the main motivations of the current study. Another was an interest in analyzing near

¹Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain.

²Saint Petersburg AIDS Center, St. Petersburg, Russia.

³Botkin's Infectious Diseases Hospital, St. Petersburg, Russia.

⁴D. I. Ivanovsky Institute of Virology, Moscow, Russia.

⁵World Health Organization, Geneva, Switzerland.

*These authors contributed equally to this work.

full-length genomes of two of the major A_{FSU} monophyletic clusters in which this variant is subdivided, previously analyzed in partial sequences, one of them designated V77I_{PR}, due the presence of the characteristic V77I substitution in protease,⁸ widely circulating in Russia (nearly half of HIV-1 infections²) and other FSU countries, and the other A_{SP1}, which represents approximately 56% of HIV-1 infections in St. Petersburg,⁹ the second largest city in Russia and one of the cities most heavily affected by the HIV-1 epidemic.¹⁰

Plasma samples used for this study were collected from 2002 through 2008 in different cities of Russia from individuals belonging to diverse risk categories. Samples were selected for near full-length genome amplification and sequencing based on analyses of protease-reverse transcriptase (PR-RT) sequences showing their clustering in the V77I_{PR} or A_{SP1} A_{FSU} clusters.^{2,9} Near full-length genomes were amplified from plasma RNA in four overlapping segments by RT-PCR and sequenced as previously described.^{11,12} Electropherograms were assembled with SeqMan (DNASTAR, Madison, WI). Sequences were aligned with MAFFT v.6.¹³ Highly polymorphic segments with uncertain alignments due to variable length polymorphisms were removed for phylogenetic analyses.

Phylogenetic sequence analyses were done via maximum likelihood using RAxML.¹⁴ The substitution model used for this analysis was the general time reversible with CAT approximation for among-sites rate heterogeneity (GTR+CAT). Genomes were analyzed for the possible presence of inter-subtype recombination by bootscanning using SimPlot v3.5.¹⁵ For this analysis, a window of 300 nucleotides was used, moving in 20 nucleotide increments, and trees were constructed using the neighbor-joining algorithm and Kimura two-parameter distances. Protein sequences were analyzed using the Gene Cutter program.⁷

Near full-length genome sequences (8.8–9 kb) were obtained in 10 samples, six of the V77I_{PR} cluster and four of the A_{SP1} cluster. Epidemiological data of these samples are shown in Table 1. All 10 newly derived sequences had intact Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, and Nef open reading frames. In bootscan analyses all were uniformly of subtype A along their genomes (results not shown). In the maximum likelihood phylogenetic analysis, in which all near full-length genome A_{FSU} sequences available at the Los Alamos HIV Sequence Database⁷ (including seven from Cyprus, presumably from Eastern European individuals residing in this country^{16,17}) were included, all newly derived sequences clustered with the

other A_{FSU} viruses (Fig. 1). Within the A_{FSU} clade, the genomes of the four viruses previously classified in the A_{SP1} cluster, based on PR-RT sequences, formed a strongly supported (100% bootstrap value) cluster, and the newly derived near full-length V77I_{PR} sequences formed another strongly supported (97% bootstrap value) cluster together with 13 previously obtained genome sequences from Uzbekistan, Kazakhstan, Russia, and Cyprus, with two of the sequences, RU_915_1038 and CY173, branching basally to all others in the cluster, which formed a subcluster supported by a 96% bootstrap value (Fig. 1). [It must be noted that monophyletic clustering of the viruses of the V77I_{PR} subvariant required removal of one of the sequences from Russia downloaded from the Los Alamos Database, PokA1Rus, a probable intrasubtype recombinant between V77I_{PR} and non-V77I_{PR} A_{FSU} viruses; the inclusion of this sequence in the phylogenetic analyses caused a decline in the bootstrap support of the V77I_{PR} cluster to 45% (results not shown).]

Within the V77I_{PR} subvariant the viruses from Uzbekistan formed a local subcluster, supporting a single introduction of this subvariant in Uzbekistan (Fig. 1). It is interesting to note that none of 10 sequences from Ukraine branched in the V77I_{PR} cluster. This was also the case when PR-RT sequences retrieved from the Los Alamos Database were analyzed (results not shown). This may be explained by the fact that the HIV-1 subtype A epidemic of Ukraine may derive from a radiation predating the emergence of the V77I_{PR} variant in 1997 in Russia.⁸ The phylogenetic tree also revealed that the three subtype A viruses from Senegal previously assigned to the A3 subsubtype¹⁸ formed a cluster fully branching within the subsubtype A1 radiation, together with the A_{FSU} and East African variants, supporting the notion that the viruses classified by some authors as an A3 subsubtype in fact represent a West African variant of the A1 subsubtype.

The common ancestry of the V77I_{PR} viruses, on the one hand, and of A_{SP1} viruses, on the other, was also supported by the presence of highly characteristic amino acid residues and nucleotides that are absent or rare in A_{FSU} viruses branching outside of the clusters mentioned. Signature amino acid residues of each cluster, here defined as those present in ≥75% viruses within the corresponding cluster and in ≤10% A_{FSU} viruses outside of them, were, for V77I_{PR}, protease 77I, RT 35I and 62V, Tat 68L, Env 158T, and Nef 151E; and for the A_{SP1} viruses Vif 73Q, Tat 86K and 73S, Env 92E, and Nef 87F

TABLE 1. EPIDEMIOLOGICAL DATA OF SAMPLES USED IN THIS STUDY

<i>Virus name</i>	<i>City of sample collection</i>	<i>Gender</i>	<i>Transmission route/ risk factor</i>	<i>Year of HIV diagnosis</i>	<i>Year of sample collection</i>
RU_915_1016	St. Petersburg	F	IDU	2001	2006
RU_915_1035	St. Petersburg	F	IDU	n.a.	2006
RU_915_1038	St. Petersburg	F	IDU	2002	2006
RU_915_1041	St. Petersburg	F	Heterosexual	n.a.	2006
RU_SP_B_049	St. Petersburg	M	IDU	n.a.	2006
RU01029	Novy Urengoy	n.a.	Heterosexual	2001	2002
RU_560_1125_JA	Novy Urengoy	M	Bisexual	2005	2005
Irkutsk_5	Irkutsk	M	IDU	2003	2007
RUA001	Odintsovo, Moscow Oblast	M	Heterosexual	2008	2008
RUA007	Pushkino, Moscow Oblast	M	Heterosexual	2008	2008

IDU, injecting drug user; n.a., not available.

FIG. 1. Maximum likelihood phylogenetic tree of HIV-1 near full-length genome sequences of the AFSU variant. Names of newly derived sequences are boxed. Only bootstrap values $\geq 90\%$ are shown. Eaf and Waf denote A1 subtype variants of East and West Africa, respectively (the West African variant is designated by some authors as subtype A3¹⁸). Countries of sample collection of subtype A viruses retrieved from the Los Alamos Database are shown before the isolate's name using the ISO two-letter country codes: AZ, Azerbaijan; BY, Belarus; CD, Democratic Republic of Congo; CY, Cyprus; GE, Georgia; KE, Kenya; KZ, Kazakhstan; SN, Senegal; RU, Russia; RW, Rwanda; UA, Ukraine; UG, Uganda; UZ, Uzbekistan.

(amino acid residues are numbered as in the NL4-3 isolate). It should be noted that one of the viruses of the V77I_{PR} cluster, RUA007, lacks the V77I substitution in protease. However it contains two additional signature RT amino acid residues of this cluster, 33I and 62V. These two residues are absent from

the two viruses branching in a basal position within the V77I_{PR} cluster, RU_915_1038 and CY173, which, however, contain the V77I substitution.

Signature nucleotides in synonymous positions of the PR-RT fragment commonly analyzed for drug resistance-associated

mutations were also similarly defined for the V77I_{PR} cluster. These are nucleotides at the third position of protease codons 38 (G) and of RT codons 66 (A), 67 (T), 149 (C), and 246 (A). For the A_{SP1} viruses, five synonymous substitutions at RT codons 40, 42, 48, 144, and 160 were previously defined as characteristic of this cluster, with the presence of ≥ 3 of these nucleotides being 100% predictive for clustering within the A_{SP1} clade.⁹

To analyze the possibility of intrasubtype recombination, genome fragments of 2 kb moving in 500 nucleotide increments were phylogenetically analyzed via maximum likelihood. No evidence of tree topology incongruence between genome fragments was found in any of the newly derived sequences (results not shown).

In summary, we have obtained 10 new HIV-1 near full-length genome sequences of A_{FSU} viruses from Russia, which almost equals the number of all previously available HIV-1 near full-length genomes from this country. The analyses of these sequences support the monophyly of two major A_{FSU} clusters previously defined in PR-RT, V77I_{PR} and A_{SP1}, which represent a large proportion of HIV-1 isolates circulating in Russia and in the city of St. Petersburg, respectively. These results may contribute to a better knowledge of the propagation of HIV-1 variants in Russia and other FSU countries and may serve as a basis for the study of the correlations of these phylogenetically defined variants with biological properties and with susceptibility to immune responses relevant for the design of vaccine immunogens intended for use in these countries.

Acknowledgments

We thank the personnel at the Genomic Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, for technical assistance in sequencing. This work was funded by WHO-UNAIDS through agreement for performance of work 2011/120936-0, which was partially funded by Office of AIDS Research, National Institutes of Health, USA. Sequences have been deposited in GenBank under accessions JQ292891–JQ292900.

Author Disclosure Statement

No competing financial interests exist.

References

- Bobkov A, Cheingsong-Popov R, Selimova L, *et al.*: An HIV type 1 epidemic among injecting drug users in the former Soviet Union caused by a homogeneous subtype A strain. *AIDS Res Hum Retroviruses* 1997;13:1195–1201.
- Vázquez de Parga E, Rakhmanova A, Pérez-Álvarez L, *et al.*: Analysis of drug resistance-associated mutations in treatment-naïve individuals infected with different genetic forms of HIV-1 circulating in countries of the former Soviet Union. *J Med Virol* 2005;77:337–344.
- 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.
- 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 countries derived from sequence analyses of preepidemically transmitted viruses. *AIDS Res Hum Retroviruses* 2007;23:1599–1604.
- Nabatov AA, Kravchenko ON, Lyulchuk MG, *et al.*: 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.
- UNAIDS: Report on the global AIDS epidemic 2010: Available at http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf.
- HIV Sequence Database: Los Alamos National Laboratory. Available at <http://www.hiv-lanl.gov>.
- 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;78:11276–11287.
- Thomson MM, Vinogradova A, Delgado E, *et al.*: Molecular epidemiology of HIV-1 in St. Petersburg, Russia: Prevalence of subtype A, former Soviet Union variant, and identification of intrasubtype subclusters. *J Acquir Immune Defic Syndr* 2009;51:332–339.
- Zagdyn Z, Kovelonov A, Semikova S, *et al.*: Spread of HIV infection in Leningrad oblast during 1999–2009. *Euro Surveill* 2011;16:121–130.
- 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.
- Katoh K and Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;9:286–298.
- Stamatakis A, Hoover P, and Rougemont J: A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008;57:758–771.
- 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.
- Kousiappa I, Van De Vijver DA, and Kostrikis LG: Near full-length genetic analysis of HIV sequences derived from Cyprus: Evidence of a highly polyphyletic and evolving infection. *AIDS Res Hum Retroviruses* 2009;25:727–740.
- Kousiappa I, Achilleos C, Hezka J, *et al.*: Molecular characterization of HIV type 1 strains from newly diagnosed patients in Cyprus (2007–2009) recovers multiple clades including unique recombinant strains and lack of transmitted drug resistance. *AIDS Res Hum Retroviruses* 2011;27:1183–1199.
- Meloni ST, Kim B, Sankalé JL, *et al.*: Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: Sub-subtype A3. *J Virol* 2004;78:12438–12445.

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