PLOS ONE

HIV-1 A1 subtype epidemic in Italy originated from Africa and Eastern Europe and shows a high frequency of transmission chains involving intravenous drug users

--Manuscript Draft--

Manuscript Number:	PONE-D-15-35239R1
Article Type:	Research Article
Full Title:	HIV-1 A1 subtype epidemic in Italy originated from Africa and Eastern Europe and shows a high frequency of transmission chains involving intravenous drug users
Short Title:	HIV-1 subtype A1 in Italy
Corresponding Author:	Alessia Lai, Ph.D. University of Milan Milan, ITALY
Keywords:	Subtype A phylogeographical reconstruction, HIV-1 A1 subtype spread in Italy, phylodynamic analysis,A1 epidemics,dated phylogeny
Abstract:	Subtype A accounts for only 12% of HIV-1 infections worldwide but predominates in Russia and Former Soviet Union countries of Eastern Europe. After an early propagation via heterosexual contacts, this variant spread explosively among intravenous drug users. A distinct A1 variant predominates in Greece and Albania, which penetrated directly from Africa. Clade A1 accounts for 12.5% of non-B subtypes in Italy, being the most frequent after F1 subtype. Aim of this study was to investigate the circulation of A1 subtype in Italy and trace its origin and diffusion through phylogenetic and phylodynamic approaches. The phylogenetic analysis of 113 A1 pol sequences included in the Italian ARCA database, indicated that 71 patients (62.8%) clustered within 5 clades. A higher probability to be detected in clusters was found for patients from Eastern Europe and Italy (88.9% and 60.4%, respectively) compared to those from Africa (20%) ($p<.001$). Higher proportions of clustering sequences were found in intravenous drug users with respect to heterosexuals (85.7% vs. 59.3%, $p=.056$) and in women with respect to men (81.4% vs. 53.2%, $p<.006$). Subtype A1 dated phylogeny indicated an East African origin around 1961. Phylogeographical reconstruction highlighted 3 significant groups. One involved East European and some Italian variants, the second encompassed some Italian and African strains, the latter included the majority of viruses carried by African and Italian subjects and all viral sequences from Albania and Greece. Subtype A1 originated in Central Africa and spread among East European countries in 1982. It entered Italy through three introduction events: directly from East Africa, from Albania and Greece, and from the area encompassing Moldavia and Ukraine. As in previously documented A1 epidemics of East European countries, HIV-1 A1 subtype spread in Italy in part through intravenous drug users. However, Eastern European women contributed to the penetration of such variant, probably through sex work.
Order of Authors:	Alessia Lai, Ph.D.
	Giorgio Bozzi
	Marco Franzetti
	Francesca Binda
	Francesco R Simonetti
	Andrea De Luca
	Valeria Micheli
	Paola Meraviglia
	Patrizia Bagnarelli
	Antonio Di Biagio
	Laura Monno

	Francesco Saladini
	Maurizio Zazzi
	Gianguglielmo Zehender
	Massimo Ciccozzi
	Claudia Balotta
Opposed Reviewers:	
Response to Reviewers:	Reviewer #1 The manuscript was revised to improve English. Regarding the difficulty in understanding the manuscript it has to be taken into account that phylogenetic studies are complex and sophisticated and need to be conducted with ad hoc biostatistical and mathematical methods. We tried when possible to clarify the performed analysis, while the usage of technical language must be applied. We, as well as other research groups, published few papers in this field on Plos One using the same language and depth of details of our experimental procedures (Zehender G, et al. PLoS One. 2012; Lai A, et al. PLoS One. 2012; Zehender G, et al. PLoS One. 2012; Zehender G, et al. PLoS One. 2013; Frentz D, et al. PLoS One. 2014). Specific comments: 1) Second paragraph (line 200 on) – no indication of what is analyzed in this paragraph. S1 figure totally unclear, including the legend. The introduction of the second paragraph, as well as the legend of S1 figure, were modified to indicate what is analyzed. 2) Phylodynamic analysis – no indication of what is analyzed in this paragraph. Figure 1 and legend are totally unclear. The introduction of the this paragraph was modified. 3) Phyleogeographical reconstruction – what is S2 figure? What should the reader understand? The tree in Fig. 2 is from 229 isolates. Should Authors explain that these isolates are from the Los Alamos Database? Do they include the ones analyzed in the present study? What is a “MCC tree”? As described in Patients and methods, regarding the Phylogenetic dataset (line 102-115), the phylogenetic signal was evaluated on phylogeographic dataset of 229 subtype A1 pol gene sequences obtained either from the Los Alamos HIV Sequence Database (n=119) (www.hiv.lanl.gov) or from Italian patients (n=37). S2 Figure represent the result of TreePuzzle program as described in Patients and methods-Likelihood mapping analysis (line 147-155). Additional informations have been added in the figure legend. The MCC tree is the maximum clade credibility tree as indicated at line 143. 4) Figure 3 – is this coming from the literature? Did the Authors generate it? If yes, how. This is totally unclear. We generated Figure 3 using SPREAD program, as reported in Patients and methods: ‘To provide a spatial projection, the migration routes indicated by the tree were visualized using Google Earth (http://earth.google.com) and the SPREAD program (available at http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html)’. The original Figure 3 was removed and replaced with a new one. 5) Last paragraph and Figure 4 – no explanation whatsoever. More details have been added on legend of Figure 4 and in the Patients and methods (Phylogenetic dataset paragraph). Reviewer #2: In this manuscript Lai et al are using a relevant number of samples to perform an in-depth analysis of the HIV-1 subtype A strain diversity in Italy. They report that, similar to other Western countries in which epidemic spread is governed by founder effect, >60% of the subtype A strains from Italy cluster in 5 subclades that can be traced to Eastern Europe. They also report that, in agreement with previously published data, sequences from European patients have a higher probability to be grouped in specific clusters than those from African patients and that a higher proportion of clustering sequences can be found in intravenous drug users than in heterosexuals and in women with respect to men. These results were expected and are in agreement with known features of molecular epidemiology of HIV-1. They also dated the origin of Subtype A in Italy and established that the strains most likely emerged from East Africa around 1961.

	Altogether, the paper is pretty straight forward and of some interest for the field of HIV diversity. Data might be built into the overall global picture of HIV diversity. We thank the reviewer for the comments and the precise revision that improve the quality of the present version of the manuscript. We addressed specific issues as follows. There are several minor issues to the manuscript. 1. First, the reference list is pretty slim. The paper is focused on the emergence of Subtype A in Italy from an easter Europe source (i.e., Moldavia/Ukraine). Yest relatively few references are listed relative to the epidemic in Ukraine and none relative to that in Moldavia. They should fix the reference list including all the relevant sources. We added the requested, as well as other references. 2. Similarly, when they make the discussion on the lower likelihood of HIV sequences from Africa to be grouped in clusters, they should provide references (this is an issue that has been discussed for a long time, starting with Vidal et al JVI 2000) This point is addressed now in the Discussion section. 3. As the sequences used for the tree construction in Figure 2 are not named, when they list the countries of origin, they should include the reference numbers linked to these sequences to permit the readership to assess the relevance. We added a supplementary table with requested information 4. I would rename the conclusion section to discussion. We renamed it. 5. In the discussion related to the possible recombinant nature of the analyzed strains, they should state that the most critical strain (CRF2) has the breaking points in the prot/pol region used for analyses and that there is thus unlikely that the included strains are Ib_NG like. This point was added in the Discussion section.
Additional Information:	
Question	**Response**
Financial Disclosure Please describe all sources of funding that have supported your work. A complete funding statement should do the following: Include **grant numbers and the URLs** of any funder's website. Use the full name, not acronyms, of funding institutions, and use initials to identify authors who received the funding. **Describe the role** of any sponsors or funders in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. If they had no role in any of the above, include this sentence at the end of your statement: "*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*" If the study was **unfunded**, provide a statement that clearly indicates this, for example: "*The author(s) received no specific funding for this work.*" * typeset	The author(s) received no specific funding for this work.
Competing Interests	The authors have declared that no competing interests exist

You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or non-financial competing interests. Do any authors of this manuscript have competing interests (as described in the PLOS Policy on Declaration and Evaluation of Competing Interests)? **If yes**, please provide details about any and all competing interests in the box below. Your response should begin with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests:* **If no** authors have any competing interests to declare, please enter this statement in the box: "*The authors have declared that no competing interests exist.*" * typeset	
Ethics Statement You must provide an ethics statement if your study involved human participants, specimens or tissue samples, or vertebrate animals, embryos or tissues. All information entered here should **also be included in the Methods section** of your manuscript. Please write "N/A" if your study does not require an ethics statement. **Human Subject Research (involved human participants and/or tissue)** All research involving human participants must have been approved by the authors' Institutional Review Board (IRB) or an equivalent committee, and all clinical investigation must have been conducted according to the principles expressed in the Declaration of Helsinki. Informed consent, written or oral, should also have	Subjects signed an informed consent to have their anonymized data stored on a central server of the ARCA database (www.dbarca.net) and used for research studies. Authors working in the clinical setting interacted with some of the patients included in the study as part of their own routine HIV care. No additional visits were scheduled, apart from those regularly planned for HIV monitoring, according to Italian national guidelines. ARCA is an observational HIV cohort approved by the Regional Ethical Committee of Tuscany (Comitato Etico Area Vasta Toscana Sudest). Epidemiological data (gender, risk category, country of origin, date of diagnosis and age) were collected by physicians from patient medical records and then included in the databases together with virological, immunological and treatment information. The study was conducted in accordance with the 1964 Declaration of Helsinki and the ethical standards of the Italian Ministry of Health.

been obtained from the participants. If no consent was given, the reason must be explained (e.g. the data were analyzed anonymously) and reported. The form of consent (written/oral), or reason for lack of consent, should be indicated in the Methods section of your manuscript. Please enter the name of the IRB or Ethics Committee that approved this study in the space below. Include the approval number and/or a statement indicating approval of this research. **Animal Research (involved vertebrate animals, embryos or tissues)** All animal work must have been conducted according to relevant national and international guidelines. If your study involved non-human primates, you must provide details regarding animal welfare and steps taken to ameliorate suffering; this is in accordance with the recommendations of the Weatherall report, "The use of non-human primates in research." The relevant guidelines followed and the committee that approved the study should be identified in the ethics statement. If anesthesia, euthanasia or any kind of animal sacrifice is part of the study, please include briefly in your statement which substances and/or methods were applied. Please enter the name of your Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board, and indicate whether they approved this research or granted a formal waiver of ethical approval. Also include an approval number if one was obtained. **Field Permit** Please indicate the name of the institution or the relevant body that granted permission.	
Data Availability PLOS journals require authors to make all data underlying the findings described in	Yes - all data are fully available without restriction

their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found. Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable.** ***If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.*** Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	
Please describe where your data may be found, writing in full sentences. **Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted.** If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details. If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, "All relevant data are within the paper and its Supporting Information files." If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, "All XXX files are available from the XXX database (accession number(s) XXX, XXX)." If this information will only be available after acceptance, please indicate this by ticking the box below. If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example: "Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data."	All relevant data are within the paper and its Supporting Information files.

“Data are from the XXX study whose authors may be contacted at XXX.” * typeset	
Additional data availability information:	

Dear Editor,

Please find enclosed a manuscript entitled "**HIV-1 A1 subtype epidemic in Italy originated from Africa and Eastern Europe and shows a high frequency of transmission chains involving intravenous drug users**" by Lai et al., which shows the features of the HIV-1 subtype A1 epidemic in Italy and traces its origin and dissemination from East Africa to Eastern European countries. In submitting this manuscript to PLoS One, we affirm that none of the work has been published, nor has it been submitted for publication elsewhere, including the Internet.

This study is the result of a nation-based collaboration supported by a comprehensive database that includes sequence, demographic and viro-immunologic data of more than 30,000 patients. We analyzed a considerable number of HIV-infected individuals carrying A1 subtype, the second most prevalent non-B clade in Italy. This study provides accurate information about the association of A1 subtype with clear geographic origin, as well as the routes of transmission among patients living in Italy and carrying such variant. Our results indicates that A1 clade entered Italy through three distinct introduction events from East Africa, Southern Balkan Peninsula and the area encompassing Moldavia and Ukraine. The present study also investigated the HIV-1 A1 epidemic worldwide, with the aim to establish the origin of the A1 variant in the European and African context. At present, only partial data regarding specific countries, such as Albania and Greece, are published on this topic.

We would like to suggest Marco Salemi of the University of Florida as an Academic Editor.

Sincerely,

Alessia Lai, PhD

Department of Biomedical and Clinical Sciences "Luigi Sacco"

University of Milan

Via G.B. Grassi 74, 20157, Milan, Italy

Alessia.lai@unimi.it

HIV-1 A1 subtype epidemic in Italy originated from Africa and Eastern Europe and shows a high frequency of transmission chains involving intravenous drug users

Alessia Lai[1]*, Giorgio Bozzi[1], Marco Franzetti[1], Francesca Binda[1], Francesco R Simonetti[1], Andrea De Luca[2], Valeria Micheli[3], Paola Meraviglia[4], Patrizia Bagnarelli[5], Antonio Di Biagio[6], Laura Monno[7], Francesco Saladini[8], Maurizio Zazzi[8], Gianguglielmo Zehender[1], Massimo Ciccozzi[9], Claudia Balotta[1]

[1]Department of Biomedical and Clinical Sciences "Luigi Sacco", Infectious Diseases and Immunopathology Section, 'L. Sacco' Hospital, University of Milan, Milan, Italy; [2]Division of Infectious Diseases, Siena University Hospital, Siena; [3]Laboratory of Microbiology, 'L. Sacco' Hospital, Milan, Italy; [4]2nd Division of Infectious Diseases, 'L. Sacco' Hospital, Milan, Italy; [5]Department of Biomedical Science, Section of Microbiology, Laboratory of Virology, University Politecnica delle Marche, Ancona, Italy; [6]Infectious Diseases Clinic, San Martino Hospital, Genoa; [7]Division of Infectious Disease, University of Bari, Bari, Italy; [8]Department of Medical Biotechnology, University of Siena, Siena, Italy; [9] Epidemiology Unit, Department of Infectious, Parasite and Immune-Mediated Diseases, Italian Institute of Health, Rome, Italy.

*Corresponding Author:

Dr. Alessia Lai, PhD: Department of Biomedical and Clinical Sciences "Luigi Sacco", Infectious Diseases and Immunopathology Section, University of Milan, Via G.B. Grassi 74, 20157 Milan, Italy – Phone (+39) 0250319775, Fax: (+39) 0250319768 – e-mail: alessia.lai@unimi.it

Abstract

Subtype A accounts for only 12% of HIV-1 infections worldwide but predominates in Russia and Former Soviet Union countries of Eastern Europe. After an early propagation via heterosexual contacts, this variant spread explosively among intravenous drug users. A distinct A1 variant predominates in Greece and Albania, which penetrated directly from Africa. Clade A1 accounts for 12.5% of non-B subtypes in Italy, being the most frequent after F1 subtype. Aim of this study was to investigate the circulation of A1 subtype in Italy and trace its origin and diffusion through phylogenetic and phylodynamic approaches. The phylogenetic analysis of 113 A1 *pol* sequences included in the Italian ARCA database, indicated that 71 patients (62.8%) clustered within 5 clades. A higher probability to be detected in clusters was found for patients from Eastern Europe and Italy (88.9% and 60.4%, respectively) compared to those from Africa (20%) (*p<.001*). Higher proportions of clustering sequences were found in intravenous drug users with respect to heterosexuals (85.7% *vs*. 59.3%, *p=.056*) and in women with respect to men (81.4% *vs.* 53.2%, *p<.006*).

Subtype A1 dated phylogeny indicated an East African origin around 1961. Phylogeographical reconstruction highlighted 3 significant groups. One involved East European and some Italian variants, the second encompassed some Italian and African strains, the latter included the majority of viruses carried by African and Italian subjects and all viral sequences from Albania and Greece.

Subtype A1 originated in Central Africa and spread among East European countries in 1982. It entered Italy through three introduction events: directly from East Africa, from Albania and Greece, and from the area encompassing Moldavia and Ukraine. As in previously documented A1 epidemics of East European countries, HIV-1 A1 subtype spread in Italy in part through intravenous drug users. However, Eastern European women contributed to the penetration of such variant, probably through sex work.

Background

The evolution of HIV-1 subtype A, similar to subtype F, gave rise to different sub-subtypes (A1, A2, A3, A4) [1,2], defined as sub-clusters and formed by distinct lineages highly related to the parental clade. The genetic distance between these subclades is about half of that between subtypes.

Subtype A is responsible of about 80% of the HIV-1 epidemic in Former Soviet Union (FSU) countries [3-5]. Evolving from a Central African ancestor, this variant initially spread through heterosexual (HE) contact in Southern Ukraine. Subtype A then begun to spread dramatically among injective drug users (IDUs) in the port city of Odessa around 1995. During the following years, the HIV-1 epidemic sustained by subtype A among IDU (labeled A_{fsu} or *IDU-A*) propagated through other FSU countries such as Russia and Estonia [6]. The explosive spread of this clade within the population of FSU IDUs was due to effective transmission networks taking place in the favorable setting of the massive socio-economic crisis which followed the fall of the Soviet Union Republics [7]. In the largely populated city of St. Petersbourg the proportion of A_{fsu} infections increased from 37% in the 1997-2004 period to 91% in 2006 of HIV-1 infections [6].

A feature of the IDU-A variant is a high homogeneity of viral sequences, probably due to the high transmission rates among IDUs following single introduction events in distinct geographic areas and the recent spread of the epidemic [8].

After its explosive diffusion in Russia, subtype A spread in neighboring countries [9-12]. In Bulgaria, subtype A was isolated only in a single individual before 1995, however, it accounted for 27% infections in further years [13], suggesting a late introduction through multiple events.

The prevalence of subtype A is approximately 2% in Western and Central Europe, however this variant has established extensive epidemics in some Mediterranean countries, such as Albania, Cyprus and Greece. In Greece, clade A1 is the most common non-B subtype (20.6%), rising from a 6% prevalence in 1984 to 42% in 2004 [14,15]. The introduction of this subtype in Greece dates back to the first epidemic phase in the country (time of the most recent common ancestor, tMRCA, 1978), probably originating from Central Africa [16]. Differently from other European countries, subtype A is the most frequent in long-dated residents

72 compared with subtype B, and was involved in sexual transmission risk groups more lately [14]. Similarly,
73 the clade A epidemic in Albania probably arose from Greece, Albanian and Greek sequences being more
74 related to African ones than to Eastern European ones [17].
75 Clade A is a parental subtype in most of known circulating recombinant forms (CRFs), particularly in the
76 most prevalent ones. These CRFs are estimated to sustain 27% HIV-1 infections globally [18], especially
77 CRF02_AG in Western Africa [19], CRF01_AE in Thailand [20] and CRF03_AB in the FSU. The high prevalence
78 of co-circulating subtype A and B in Eastern Europe represented the background for the origin of CRF03_AB
79 in Southern Ukraine, giving rise to an outbreak through IDUs in the city of Kaliningrad in 1996 [21].
80 Due to migration fluxes from Africa, FSU and South America, non-B subtype circulation is increasing in Italy
81 as well as in all Western countries of Europe. The overall prevalence of infections due to non-B clade in Italy
82 is 11.4% having raised from 2.6% to 18.9% over three decades. Among these, subtype A is the second in
83 prevalence (12.7%) after clade F1 (23.7%) [22]. The aim of this study was to investigate the features of A1
84 subtype circulation in Italy and trace its origin and diffusion through phylogenetic and phylodynamic
85 approaches.
86

Patients and methods

Study population

We studied 113 individuals carrying HIV-1 A1 subtype. Patients were sampled from 1999 through 2011. Subjects signed an informed consent to have their anonymized data stored on a central server of the ARCA database (www.dbarca.net) and used for research studies. Authors working in the clinical setting interacted with some of the patients included in the study as part of their own routine HIV care. No additional visits were scheduled, apart from those regularly planned for HIV monitoring, according to Italian national guidelines. ARCA is an observational HIV cohort approved by the Regional Ethical Committee of Tuscany (Comitato Etico Area Vasta Toscana Sudest). HIV-1 protease (PR) and partial reverse transcriptase (RT) sequences were generated by local centers for routine drug resistance testing at diagnosis or prior to the start of antiretroviral therapy or at treatment failure. Epidemiological data (gender, risk category, country of origin, date of diagnosis and age) were collected by physicians from patient medical records and then included in the databases together with virological, immunological and treatment information. Only the earliest available HIV-1 genotype was considered for each patient. The study was conducted in accordance with the 1964 Declaration of Helsinki and the ethical standards of the Italian Ministry of Health.

Phylogenetic dataset

The analysis of epidemiological clusters was performed on the starting dataset of 113 HIV-1 A1 subtypes.
The evolutionary population dynamics was estimated in the dataset of Italian patients (n=53).
A reference set of 229 subtype A1 *pol* gene sequences was obtained from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov). Viral sequences were selected according to the following inclusion criteria: i) sequences had already been published in peer-reviewed journals; ii) the subtype assignment of each sequences classified as non-recombinant was certain; iii) the city/state of origin and the year of sampling were known and clearly established in the original publication. The final reference set includes subtype A1 sequences of Central Africa (n=22; Democratic Republic of Congo=13, Central African Republic=3,

Gabon=6), East Africa (n=52; Uganda=17, Tanzania=16, Kenya=19), Estonia (n=9), Latvia (n=19), Lithuania (n=5), Moldavia (n=5), Ukraine (n=20), Byelorussia (n=10), Russia (n=20), Kazakhstan (n=6), Uzbekistan (n=7), Greece (n=9), Albania (n=8). The accession number list of phylogeographic references is provided in S1 Table Thirty seven randomized sequences belonging to patients with established Italian citizenship present in a starting dataset were included. The phylogeographic dataset of 229 isolates was used to evaluate the migration pattern of HIV-1 subtype A1 circulation.The phylogenetic signal was evaluated on the phylogeographic dataset.

Alignment and phylodynamic analysis

The sequences were manually aligned using Bioedit software (v7.2.5; http://www.mbio.ncsu.edu/bioedit/bioedit.html). All sites of major antiretroviral drug resistance mutations in the protease and reverse transcriptase regions, identified through the last updated Surveillance Drug Resistance Mutations (SDRM) (http://hivdb.stanford.edu/cgi-bin/AgMutPrev.cgi) and International AIDS Society (IAS) (http://www.iasusa.org/guidelines/index.html) lists for naïve and treated patients respectively, were removed to avoid bias due to convergence related to drug resistance. The final alignment encompassed 834 nucleotides.

The Bayesian phylogenetic tree was reconstructed by means of both MrBayes and Beast using the general time reversible (GTR)+ gamma distribution (G)+ proportion of invariable sites (I) model, which was selected using the Modeltest v. 3.7 program. A Markov Chain Monte Carlo (MCMC) search was made for $10x10^6$ generations using tree sampling every 100th generation and a burn-in fraction of 50%. Statistical support for specific clades was obtained by calculating the posterior probability (pp) of each monophyletic clade, and a posterior consensus tree was generated after a 50% burn-in. Clades with a pp of 1 were considered epidemiologically related clusters.

Dated trees, evolutionary rates and population growth were co-estimated by using Beast v.1.8.0 [23] software (http://beast.bio.ed.ac.uk) using the previous model. Different coalescent priors (constant

population size, exponential growth, logistic growth and Bayesian skyline plot) were tested using both strict and relaxed molecular clock models. Different clock models were compared by calculating the Bayes Factor (BF) [15]. The difference (in $\log_e$ space) of marginal likelihood between any two models is $\log_e$ BF. Evidence against the null model (i.e. the one with lower marginal likelihood) was indicated by: 6>(2 $\log_e$ BF)>2 (positive), 10>(2 $\log_e$ BF)>6 (strong) and (2 $\log_e$ BF)>10 (very strong) [16]. BF calculations were performed with Tracer v.1.5 software (http://beast.bio.ed.ac.uk/Tracer). The BF analysis showed that the relaxed lognormal molecular clock fitted the data better than the strict clock model (2lnBF=114.5). Chains were run for 50 million generations with sampling every 5,000 generations. The results were visualized in Tracer v.1.5 and uncertainty in the estimates was indicated by 95% highest probability density (HPD) intervals. Convergence was assessed based on the ESS (effective-sample size) value and only parameter estimates with ESS>300 were accepted. The maximum clade credibility (MCC) tree was then selected from the posterior tree distribution using TreeAnnotator v.1.8.0 available in the BEAST software package. Final trees were visualized and manipulated with FigTree v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Likelihood mapping analysis

In order to obtain an overall feature of the phylogenetic signal present in the HIV-1 A1 sequences, we made a likelihood-mapping analysis of 10,000 random quartets generated using TreePuzzle (Schmidt et al., 2002). A likelihood map consists of an equilateral triangle: each dot within the triangle represents the likelihoods of the three possible unrooted trees for a set of four sequences (quartets), randomly selected from the dataset. The dots close to the corners or at the sides respectively represent tree-like (fully resolved phylogenies in which one tree is clearly better than the others) or network-like phylogenetic signals (three regions for which it is not possible to decide between two topologies). The central area of the map represents a star-like signal (the region where the star tree is the optimal tree).

Phylogeographic analysis

Analyses were performed using Bayesian Skyline coalescent tree prior with a relaxed clock (uncorrelated Lognormal model). Chains were run for 100 million generations with sampling every 10,000. The results were visualized using Tracer v.1.5. The ESS value for each parameter was >300, indicating sufficient mixing of the Markov chain. Rates with a BF of >3 were considered well supported and formed the migration pathway. The MCC tree was selected from the posterior tree distribution after a 10% burn-in using the TreeAnnotator program, version 1.8.0. The final tree was visualized using FigTree version 1.4.0. To provide a spatial projection, the migration routes indicated by the tree were visualized using Google Earth (http://earth.google.com) and the SPREAD program (available at http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html).

Viral Gene Flow Analysis

The MacClade program version 4 (Sinauer Associates, Sunderland, MA) was used to test viral gene out/in using a modified version of the Slatkin and Maddison test [24]. A one-character data matrix was obtained from the original dataset by assigning a one-letter code, indicating the geographical origin, to each taxon in the tree. The putative origin of each ancestral sequence (i.e., internal node) in the tree was then inferred by finding the most parsimonious reconstruction (MPR) of the ancestral character using either the ACCTRAN or DELTRAN option. The final tree length, that is the number of observed viral gene flow events in the genealogy, is computed and compared to the tree-length distribution of 10,000 trees obtained by random joining-splitting (null distribution). Observed genealogies significantly shorter than random trees indicate the presence of subdivided populations with restricted gene flow. The viral gene flow (migrations) was traced with the 'State changes and stasis tool' (MacClade software), which counts the number of changes in a tree for each pairwise character state. Gene flow was also calculated for the null distribution to assess whether the gene flow events observed in the actual tree were significantly higher (>95%) or lower (<95%) than the values in the null distribution at the $p = .05$ level.

187

188 **Statistical methods**

189 Categorical variables were summarized as frequencies and comparisons between groups were performed
190 using the $\chi 2$ or the Fisher exact test. For all analyses an α error of 5% was considered. The crude and
191 Mantel–Haenszel-adjusted odds ratios of inclusion in clusters with 95% CIs were calculated in univariate
192 and multivariate analyses. Analyses were performed using the SPSS software package (v.21.0, SPSS Inc.
193 Chicago, IL, USA).

194

Results

Patient characteristics

One-hundred thirteen HIV-1 A1 infected patients with sequence data spanning a 13-year period (Table 1) were included. Italians, East Europeans, Africans, South Americans and Asians accounted for 46.9% (n=53), 15.9% (n=18), 8.8% (n=10), 1.8% (n=2) and 0.1% (n=1) of patients with known country of origin, respectively. Known risk factors were distributed as follows: 64.3% (n=27) HEs, 19% (n=8) men having sex with men and 16.7% (n=7) IDUs. Males accounted for 59% of patients (n=62). Median age was 35 years (range: 24-44 years). Naïve patients accounted for 21.2% (n=24); only two patients harbored resistant strains (8.3%) and only mutations selected by Protease Inhibitors (PI) were present.

Associations between subtype A and patients characteristics

The phylogenetic trees were constructed on 113 individuals carrying HIV-1 A1 subtype. No differences were observed in the topology of trees obtained using MrBayes or BEAST programs. The BEAST tree of HIV-1 A1 sequences is shown in S1 Fig. Based on the latter tree we identified 5 highly supported clades (with posterior probability, pp, equal to 1) (S1 Fig) that included 71 patients (62.8%). A large network (#5) encompassed 59 isolates (83.1% of clustering sequences), while the remaining four included three patients each. The major clade involved all patients from Eastern Europe and included 47.2% of Italians (n=25). Noteworthy, female gender accounted for 54.2% (n=32) in cluster #5. Differently to the other networks, cluster #4 involved only Italian subjects. Clusters #1 and #4 included only treated subjects. None of the clusters included exclusively subjects with the same gender.

The overall epidemiological characteristics and those of patients included or not within epidemiological networks are described in Table 1.

Patients from Eastern Europe and Italy had a higher probability to be detected in clusters (n=16, 88.9% and n=32, 60.4%, respectively) compared to those from Africa (n=2, 20%) ($p<.001$). The proportion of clustering sequences was slightly higher in IDUs (n=6, 85.7%) and HEs (n=16, 59.3%) compared to MSM (n=2, 25%)

(*p=.056*). Women carried a higher proportion of clustering sequences (n=35, 81.4%) compared to men (n=33, 53.2%) (*p<.006*). Among clustering females with known risk factor, 80% (n=8) were HEs and 20% IDUs. Seven of HE females were from East Europe (87.5%).

Male gender was predominant both in total population and in Italian patients within clusters (n=38, 71.7% *vs.* n=19, 59.4%, respectively) (*p<.001*), whereas female gender was prevalent among patients from Eastern Europe (n=13, 16.7% *vs.* n=15, 18.7%) (*p=.029*). No difference was observed in the distribution of naïve subjects inside or outside clusters, however 19 (79%) naïve individuals were present in the epidemiological chains. Patients involved in such chains were significantly younger (33 years, range: 28.5-40 years) than patients outside clusters (39 years, range: 34-53.5 years) *(p=.003)*.

Phylodynamic analysis

The phylodynamic analysis was conducted on the dataset of patients with certain Italian citizenship (n=53) carrying HIV-1 A1 subtype.The Bayesian Skyline plot showed the changes in population size at different times from the root of the tree to the time of the most recent isolates (year 2011). The mean substitution rate for the analyzed dataset was 1.11×10^{-3} (95% HPD: 2.82×10^{-4} – 1.87×10^{-3}). The effective number of infections grew exponentially approximately between 1990 and 1998 and reached a plateau around 2000, when the effective number of infections remained constant until the last calendar years (Fig 1).

Phylogeographical reconstruction

The likelihood mapping analysis of the HIV1-A1 dataset of 229 isolates showed that the percentage of dots falling in the central area of the triangles was 4.2%, thus indicating a fully resolved phylogenetic signal (S2 Fig).

The phylogeographic maximum clade credibility (MCC) tree, obtained grouping the 229 isolates within ten discrete groups on the basis of the sampling location (IT, Italy; CA, Central Africa; EA, East Africa; EE,

244 Estonia; LL, Latvia/Lithuania; MU, Moldavia/Ukraine; BY, Byelorussia; RU, Russia; KU,
245 Kazakhstan/Uzbekistan; GA, Greece/Albania) is shown in Fig 2.
246 The most probable location of the root of the tree was East Africa, supported by a state posterior
247 probability (spp) significantly higher than that of the other locations (spp=0.58 *vs.* spp=0.42 for Central
248 Africa, the second most probable location of the root). The date of the most recent common ancestor
249 (MRCA), corresponding to the East African root, was estimated to be 51 years (95% HPD=42.1-69.4 years),
250 corresponding to the year 1960.
251 Phylogeographical reconstruction highlighted 3 highly supported groups (A, pp=1; B, pp=0.97; and C, pp=1).
252 Group A involved all sequences from Eastern Europe and the most probable location is Byelorussia
253 (spp=0.76). Sixteen Italian sequences were interspersed within this group. The mean tMRCA of the A group
254 was 29.7 years, (95% HPD=22-35.7 years) corresponding to 1981. In this group five clusters can be
255 identified (1a, pp=1; 2a, pp=1; 3a, pp=1; 4a, pp=1; and 5a, pp=0.98). The first cluster (1a) included 27
256 isolates (10 from Kazakhstan/Uzbekistan, 3 from Estonia, 6 from Russia and 8 from Latvia/Lithuania);
257 cluster 2a involved 12 strains (7 from Russia, 4 from Estonia and 1 from Italy); cluster 3a contained only
258 sequences from Latvia/Lithuania (n=13); cluster 4a contained 7 isolates (3 from Latvia/Lithuania, 2 from
259 Russia and 2 from Italy); cluster 5a included 7 isolates (4 from Italy and 3 from Moldavia/Ukraine). The
260 mean tMRCA of these clusters range from 15-19 years (95% HPD = 12-20 years).
261 Group B involved 9 African sequences and two Italian strains and the most probable location is East Africa
262 (spp=1). The mean tMRCA of B group was 41.8 years (95% HPD =32.8-42.7 years).
263 Group C involved the majority of African sequences (spp=1 for East Africa), 18 Italian sequences, all strains
264 from Greece/Albania and 2 Central African sequences. The mean tMRCA of group C was 44 years, (95% HPD
265 =36.1-47.8 years) corresponding to 1967. In this group three clusters were identified (1c, pp=0.99; 2c,
266 pp=0.99; and 3c, pp=1). The first cluster (1c) included 16 isolates (11 from East Africa, 4 from Italy and 1
267 from Central Africa); the cluster 2c involved 27 strains (19 from East Africa, 5 from Italy and 1 from Central
268 Africa); cluster 3c contained 30 sequences (6 from East Africa, 15 from Greece/Albania and 9 from Italy).

The mean tMRCA of these clusters range from 38-42 years (95% HPD = 32.2-44 years). tMRCA values, confidence intervals and the locations of the main groups and clusters are listed in Table 2.

Significant Linkages

Bayesian phylogeographic analysis of 229 HIV-1 A1 isolates showed eleven non-zero rates between ten discrete locations (Italy, East Africa, Central Africa, Estonia, Latvia/Lithuania, Moldavia/Ukraine, Byelorussia, Russia, Kazakhstan/Uzbekistan and Greece/Albania). Significant linkage (BF>3.0) was found between Byelorussia and Moldavia/Ukraine (BF=6), Central Africa (BF=7.1) and Russia (BF=3.2). Russia was significantly linked with Estonia (BF=17), Kazakhstan/Uzbekistan (BF=20), Latvia/Lithuania (BF=32778.4) andMoldavia/Ukraine (BF=26.8). Italy was significantly linked with three locations: East Africa (BF=297.1), Greece/Albania and Moldavia/Ukraine (BF=32778.4 for both locations). Central Africa and East Africa are linked with a BF=5660.

The pathways of HIV-1 A1 dispersion are showed in Fig 3. HIV-1 A1 subtype originated in East Africa around 1961 and spread to Central Africa in 1972. Subsequently, the virus spread in East Europe starting from Byelorussia where the virus was introduced in 1982 from Central Africa. The virus spread to the Kazakhstan/Uzbekistan from Russia around 1996.

The HIV-1 A1 subtype reached Italy primarily from East Africa (year 1983) and subsequently from Greece/Albania and Moldavia/Ukraine around 1993-1999. A probable link related East Africa and Greece/Albania where the virus was introduced in 1983, even though a not significant BF could be observed.

Migration Rates and Fluxes

The migration pattern of HIV-1 subtype A1 circulation was calculated using the phylogeographic dataset of 229 isolates. The gene flow (migration) between different geographic areas was investigated with a modified version of the Slatkin and Maddison method. The null hypothesis of panmixia (i.e. no population

294 subdivision or complete intermixing of sequences from different geographical areas) was tested using a
295 bubblegram. The migration flow among the five locations was then estimated as observed migration in the
296 genealogy (Fig 4). The null hypothesis of panmixia was ejected by the randomization test ($p = .0001$) for all
297 the countries in the bubblegram. ACCTRAN resolving option was used for uncertainties present in the
298 resulting tree.
299 HIV-1 A1 gene flows were observed from East Africa to Central Africa (12.8%), Greece/Albania (2.6%) and
300 Italy (20.5%). Italy received migrations also from Byelorussia (5.1%), Greece/Albania (10.3%),
301 Moldavia/Ukraine (17.9%) and Russia (2.6%).
302 The gene flow from Central Africa to Byelorussia, from Byelorussia to Russia and from Moldavia/Ukraine to
303 Estonia accounted for 2.6%, 5.1% and 2.6% of the observed migrations, respectively (Fig 4). HIV-1 subtype
304 A1 gene flow indicated Russia as the epicenter of HIV-1 A1 epidemic in East Europe, branching out to
305 Byelorussia (2.6%), Moldavia/Ukraine (5.1%) and Latvia/Lithuania (7.7%) (Fig 4).
306

Discussion

The prevalence of non-B subtypes has increased over time in Italy as well as in Europe and this rise can be attributed mainly to an increasing frequency of F1, C, CRF02_AG and A1 subtypes [22].

The present study is the first detailed investigation of the HIV-1 A1 molecular epidemiology in Italy using a national database and a Bayesian phylogenetic approach.

An extraordinary proportion of variants were associated to A1 epidemiological networks (more than 62% of studied sequences). We observed a relevant role of heterosexual contacts and injecting drug use. Italian male heterosexuals and IDUs predominated while foreign-born patients were mainly heterosexual females from East Europe. The high frequency of Italian heterosexual males and Eastern Europe females within epidemiological networks indicated that sexual contacts have led to a relevant circulation of subtype A1 in Italy. Only one cluster involving patients with Italian citizenship was identified, while the majority of Italian strains were intermixed within East European patients. Additionally, despite the reported decrease of infections due to injecting drug usage (http://www.iss.it/binary/ccoa/cont/ONLINE_6_COA_2014.pdf), the circulation of A1 subtype partially remains related to this modality of infection as in previously documented A1 epidemics of East European countries [21]. In contrast with previous published data regarding other non-B subtypes circulating in Italy, such as subtype C and F1 which mobility is associated with men having sex with men, the A1 subtype remains related to heterosexuals and IDUs [25,26].

The present study also investigated the HIV-1 A1 epidemics in Italy and worldwide, with the aim to establish the origin of the A1 variant circulating in Italy as well as in the European and African context. The present phylogenetic analysis showed that subtype A1 originated, as expected, in East Africa (where subtype A is the most prevalent clade) with a recent epidemic, originating around 1961 [27].

Global subtype A1 dispersion showed a strong geographic compartmentalization in many clusters. Dispersal routes of subtype A1 from Africa to Europe were probably associated with immigration from Africa while

331 viral mobility within Eastern Europe is probably due to connections among IDUs. After its origin in East
332 Africa subtype A1 migrated in Central Africa; both variants from East and Central Africa likely originated the
333 epidemic in Europe. The spread among East European countries started from 1982, corresponding to the
334 identification of the first outbreaks in countries formerly part of Soviet Union as reported by literature [7].

335 In general, several subtype A migration pathways exist among European countries which are mainly
336 associated with heterosexuals and IDUs in Eastern Europe. Greece and Albania are a consistent exception
337 due to the establishment of a large HIV-1 subtype A local epidemic sustaining 20.6% and 56.1% of
338 infections, respectively [13,14]. In agreement with data of Paraskevis et al. [14], our findings showed a
339 monophyletic origin of subtype A epidemic in Greece and Albania and confirmed its origin in East Africa
340 (Kenya and Tanzania) with a date of origin around 1983.

341 Phylogeographical reconstruction suggests that A1 clade entered Italy through three distinct introduction
342 events: an early first getaway took place directly from East Africa, where subtype A has a high prevalence,
343 while two additional events occurred from the Southern Balkan Peninsula and the area encompassing
344 Moldavia and Ukraine. This phenomenon could be related to Eastern European women who contributed to
345 the penetration of such variant probably through sex work as demonstrated by the high proportion of
346 women in clusters. It has been documented that, during the past 10 years, Albanian female sex workers
347 traveled to other Balkan countries and to Western Europe, especially Italy and Greece. The very low
348 percentage of African subjects within epidemiological networks suggests a minor role of the African A1
349 variant compared to the East European one in the circulation of this subtype in Italy. This is in agreement
350 with the high degree of homogeneity of the virus populations carried by IDUs in the former Soviet Union
351 countries which suggests that they shares a quite recent common ancestor, as a result of a single
352 introduction. In contrast, HIV-1 A1 African sequences presented a high genetic variability, that increased in
353 Democratic Republic of Congo in comparison to in West or East Africa [11,16,28].
354 The estimation of the demographic history of subtype A1 in Italy revealed that the epidemic growth during
355 the 1990s. The stabilization of the epidemic around 2000 can be probably explained by the reduced

number of infections related to injecting drug usage and the diffusion of combination antiretroviral therapy in infected IDUs.

Globally, we hypothesize three different coeval introduction events of the A1 variant from Africa into Europe; the former involving Byelorussia [29], the second Greece and the latter Italy. The migration fluxes confirmed the results of phylogeographic reconstruction giving a strong support to the link between East Africa and Greece.

Some possible limits and biases of the present study have to be mentioned. First, although our dataset includes sequences from a national database we cannot exclude the possibility that the study population is not truly representative of all HIV-1 A1 subtype cases in Italy. Another potential limit of our study is that we analyzed a portion of *pol* gene, rather than full genome sequences, which could underestimate recombinant strains that frequently include A1 subtype portions (such as CRF01_AE or CRF02_AG, very frequent in Italy). Nevertheless, phylogenetic and recombination analysis were performed to confirm the subtype assignment on sequences classified as A1 subtype in the ARCA database by also comparing these strains with other known CRFs. Additionally, the relative lack of country of origin and risk factor information for some patients could have weakened the strength of the observed associations.

In conclusion, this study provides novel information about the A1 subtype epidemic and its geographic origin, as well as the routes of transmission among patients living in Italy and carrying such variant. Due to its founder effect, subtype B predominates the HIV-1 Italian epidemic. However, the geographic position of Italy in the center of Mediterranean basin and the increase of population mobility in Europe, had the potential to modify the previous HIV-1 epidemiological structure. The findings in the present study contribute to the understanding of this changing pattern ad highlight the importance of an extensive and continuous monitoring of the spread of HIV-1 variants in Europe.

FIGURE LEGENDS

Fig 1. Skyline plot obtained by analyzing the data set of Italian patients (n=53). Ordinate: the number of effective infections at time t (Ne(t)); abscissa: time (in years before the present). The thick solid line represents the median value and the grey area the 95% HPD of the Ne(t) estimates. The vertical lines indicate the 95% lower HPD (dotted) and the mean tMRCA estimate (bold) of the tree root.

Fig 2. Bayesian phylogeographical tree of 229 HIV-1 A *pol* sequences with branches colored on the basis of the most probable location of the descendent nodes. The correspondences between the locations and colors are shown in the panel (left), and the posterior probabilities >0.8 are indicated on the internal nodes of the tree. The scale axis below the tree shows the number of years before the last sampling time.

Fig 3. Significant non-zero HIV-1 A1 migration rates worldwide. Only the rates supported by a BF of >3 are shown in red, yellow indicates the probable rate. The relative strength of the statistical support is indicated by the color of the lines (from dark red , *id est* weak to light red *id est* strong). The map was reconstructed using SPREAD program. This figure is similar but not identical to the original image, and is therefore for illustrative purposes only.

Fig 4. Migration pattern of HIV-1 subtype A1 circulation based on phylogeographic dataset. The bubblegram shows the frequency of gene flow (migrations) to/from different geographic areas. The surface of each circle is proportional to the percentage of observed migrations in the Maximum Likelihood genealogy. Migrations were inferred with a modified version of the Slatkin and Maddison algorithm. BY, Byelorussia; CA, Central Africa; EA, East Africa; EE, Estonia; GA, Greece/Albania; IT, Italy; KU, Kazakhstan/Uzbekistan; LL, Latvia/Lithuania; MU, Moldavia/Ukraine; RU, Russia.

403 Table 1. Characteristics of 113 patients carrying subtype A1

404

		All patients	Patients within clusters	Patient outside clusters	*p*
ORIGIN % (n)	**Italy**	46.9 (53)	60.4 (32)	39.6 (21)	
	East Europe	15.9 (18)	88.9 (16)	11.1 (2)	
	Africa	8.8 (10)	20.0 (2)	80.0 (8)	*p<.001*
	South America	1.8 (2)	100 (2)	0.0 (0)	
	Asia	0.1 (1)	100 (1)	0.0 (0)	
	Unknown	26.5 (29)	96.6 (28)	3,4 (1)	
RISK FACTOR % (n)	**HEs**	64,3 (27)	59.3 (16)	40.7 (11)	
	MSM	19.0 (8)	25.0 (2)	75.0 (6)	*p=.056*
	IDUs	16,7 (7)	85.7 (6)	14.3 (1)	
GENDER % (n)	**Male**	59.0 (62)	53.2 (33)	46.8 (29)	*p<.006*
	Female	38.2 (97)	81.4 (35)	18.6 (62)	
AGE (median years)		35	33	39	

405

406

Table 2. tMRCAs, credibility intervals (95%HPD) and location of the main groups (A, B, C) and clades (1a, 2a, 3a, 4a, 5a, 1c, 2c, 3c) on the basis of the dataset of 229 HIV-1 subtype A1 isolates.

	tMRCA	**95% HPD**	**Location**
Root	51	42.1-69.4	East Africa
A	30	21.9-35.7	Byelorussia
1a	16	15.4-20	Russia
2a	19	14.2-18.4	Russia
3a	15	13.8-18.6	Latvia/Lithuania
4a	16	dic-17	Russia
5a	16	14.4-18.3	Moldavia/Ukraine
B	42	32.8-42.7	East Africa
C	44	36.1-47.8	Central Africa
1c	42	32.8-42.7	East Africa
2c	39	34-44	East Africa
3c	39	32.2-42.4	East Africa

tMRCA: time of the most recent common ancestor; HPD: highest posterior density.

ACKNOWLEDGMENTS

The authors thank all patients included in the study.

SUPPORTING INFORMATIONS

S1 Table. Accession numbers of used references.

S1 Fig. Bayesian tree of 113 HIV-1 A1 isolates obtained with Beast program. The 5 supported clusters are highlighted in red color. Node labels indicate posterior probability values (pp). Only values higher than 0.8 are shown.

S2 Fig. Likelihood mapping of the HIV-1 A1 data sets of 229 nucleotide sequences obtain using TreePuzzle program. The three corners represent fully resolved tree topologies, id est the presence of a tree-like phylogenetic signal in the given data set.

425 **References**

426 1. Gao F, Vidal N, Li Y, Trask SA, Chen Y, et al. (2001) Evidence of two distinct subsubtypes within the HIV-1
427 subtype A radiation. AIDS Res Hum Retroviruses 17: 675-688.

428 2. Meloni ST, Kim B, Sankale JL, Hamel DJ, Tovanabutra S, et al. (2004) Distinct human immunodeficiency
429 virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. J Virol 78: 12438-12445.

430 3. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1
431 during 2000-2007. AIDS 25: 679-689.

432 4. Galkin AN, Bychenko AB, Kuznecovs G, Filinova EY (2012) Isolation, characterization, and sequence
433 analysis of a novel HIV type 1 subtype A1 strain from Russia. AIDS Res Hum Retroviruses 28: 1139-
434 1146.

435 5. Smolskaya T, Liitsola K, Zetterberg V, Golovanova E, Kevlova N, et al. (2006) HIV epidemiology in the
436 Northwestern Federal District of Russia: dominance of HIV type 1 subtype A. AIDS Res Hum
437 Retroviruses 22: 1074-1080.

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

441 7. Thomson MM, de Parga EV, Vinogradova A, Sierra M, Yakovlev A, et al. (2007) New insights into the
442 origin of the HIV type 1 subtype A epidemic in former Soviet Union's countries derived from
443 sequence analyses of preepidemically transmitted viruses. AIDS Res Hum Retroviruses 23: 1599-
444 1604.

445 8. Lukashov VV, Huismans R, Rakhmanova AG, Lisitsina ZN, Akhtyrskaya NA, et al. (1999) Circulation of
446 subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St.
447 Petersburg, Russia, correlates with geographical origin of infections. AIDS Res Hum Retroviruses 15:
448 1577-1583.

449 9. Dehne KL, Pokrovskiy V, Kobyshcha Y, Schwartlander B (2000) Update on the epidemics of HIV and other
450 sexually transmitted infections in the newly independent states of the former Soviet Union. Aids 14
451 Suppl 3: S75-84.
452 10. Laga V, Vasilyev A, Lapovok I, Grigoryan S, Papoyan A, et al. (2015) HIV Type 1 Subtype A1 Dominates in
453 Armenia. Curr HIV Res 13: 219-225.
454 11. Pandrea I, Descamps D, Collin G, Robertson DL, Damond F, et al. (2001) HIV type 1 genetic diversity and
455 genotypic drug susceptibility in the Republic of Moldova. AIDS Res Hum Retroviruses 17: 1297-
456 1304.
457 12. Grossman Z, Avidor B, Mor Z, Chowers M, Levy I, et al. (2015) A Population-Structured HIV Epidemic in
458 Israel: Roles of Risk and Ethnicity. PLoS One 10: e0135061.
459 13. Salemi M, Goodenow MM, Montieri S, de Oliveira T, Santoro MM, et al. (2008) The HIV type 1 epidemic
460 in Bulgaria involves multiple subtypes and is sustained by continuous viral inflow from West and
461 East European countries. AIDS Res Hum Retroviruses 24: 771-779.
462 14. Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Paparizos V, et al. (2007) Increasing prevalence of
463 HIV-1 subtype A in Greece: estimating epidemic history and origin. J Infect Dis 196: 1167-1176.
464 15. Antoniadou ZA, Kousiappa I, Skoura L, Pilalas D, Metallidis S, et al. (2014) Short communication:
465 molecular epidemiology of HIV type 1 infection in northern Greece (2009-2010): evidence of a
466 transmission cluster of HIV type 1 subtype A1 drug-resistant strains among men who have sex with
467 men. AIDS Res Hum Retroviruses 30: 225-232.
468 16. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, et al. (2000) Unprecedented degree of
469 human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic
470 of Congo suggests that the HIV-1 pandemic originated in Central Africa. J Virol 74: 10498-10507.
471 17. Ciccozzi M, Gori C, Boros S, Ruiz-Alvarez MJ, Harxhi A, et al. (2005) Molecular diversity of HIV in Albania.
472 J Infect Dis 192: 475-479.

18. Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J (2002) Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J Acquir Immune Defic Syndr 29: 184-190.

19. Andersson S, Norrgren H, Dias F, Biberfeld G, Albert J (1999) Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from guinea-bissau with single or dual infections: predominance of a distinct HIV-1 subtype A/G recombinant in West Africa. Virology 262: 312-320.

20. Phuphuakrat A, Pasomsub E, Kiertiburanakul S, Chantratita W, Sungkanuparph S (2012) HIV type 1 integrase polymorphisms in treatment-naive and treatment-experienced HIV type 1-infected patients in Thailand where HIV type 1 subtype A/E predominates. AIDS Res Hum Retroviruses 28: 937-943.

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

22. Lai A, Riva C, Marconi A, Balestrieri M, Razzolini F, et al. (2010) Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV Med 11: 593-602.

23. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

24. Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123: 603-613.

25. Lai A, Bozzi G, Franzetti M, Binda F, Simonetti FR, et al. (2014) Phylogenetic analysis provides evidence of interactions between Italian heterosexual and South American homosexual males as the main source of national HIV-1 subtype C epidemics. J Med Virol 86: 729-736.

26. Lai A, Simonetti FR, Zehender G, De Luca A, Micheli V, et al. (2012) HIV-1 subtype F1 epidemiological networks among Italian heterosexual males are associated with introduction events from South America. PLoS One 7: e42223.

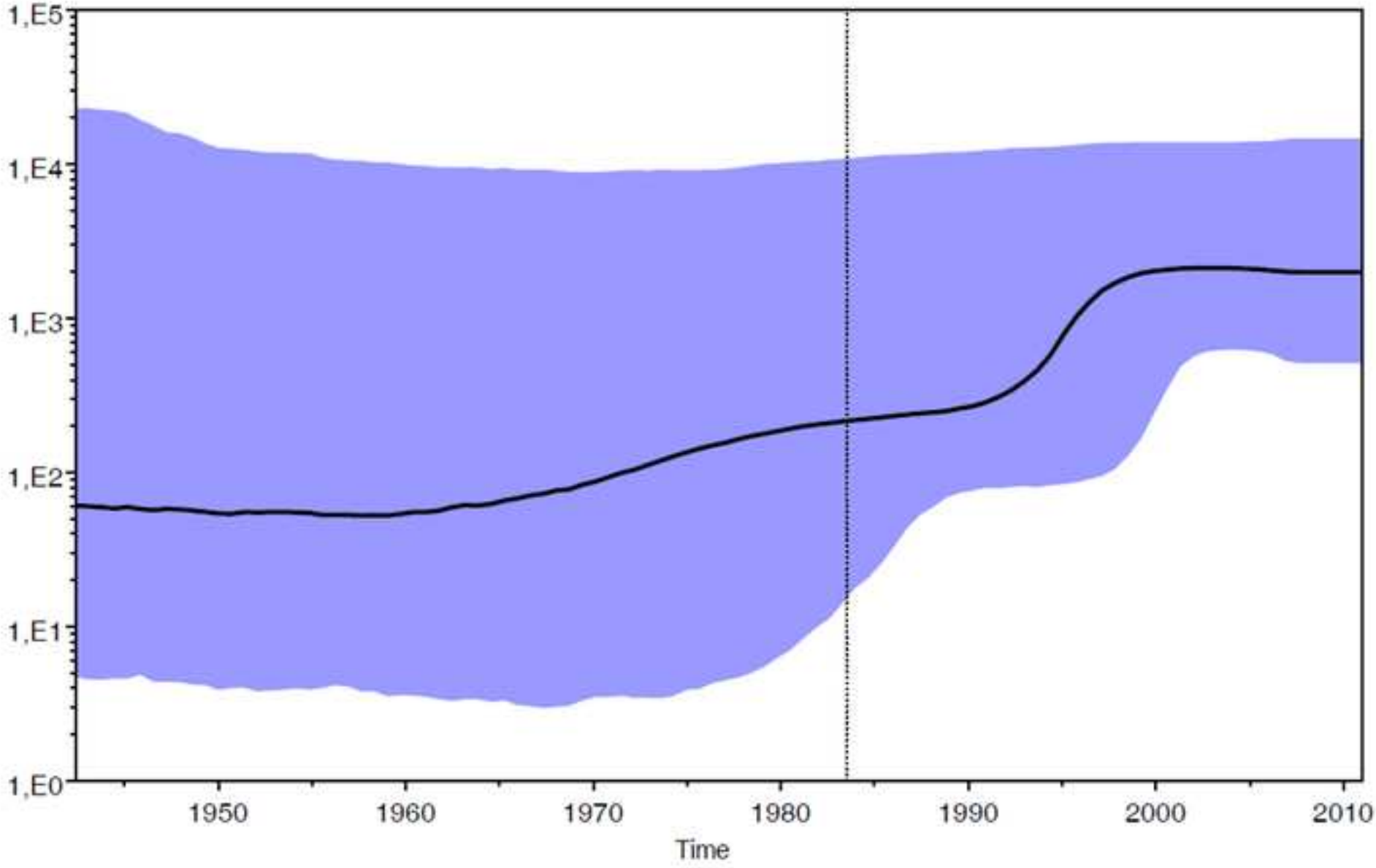
27. Njai HF, Ewings FM, Lyimo E, Foulongne V, Ngerageza D, et al. (2013) Deciphering the complex distribution of human immunodeficiency virus type 1 subtypes among different cohorts in Northern Tanzania. PLoS One 8: e81848.

28. Nyamache AK, Muigai AW, Khamadi SA (2013) Circulating trends of non-B HIV type 1 subtypes among Kenyan individuals. AIDS Res Hum Retroviruses 29: 400-403.

29. Nabatov AA, Kravchenko ON, Lyulchuk MG, Shcherbinskaya AM, Lukashov VV (2002) 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 18: 891-895.

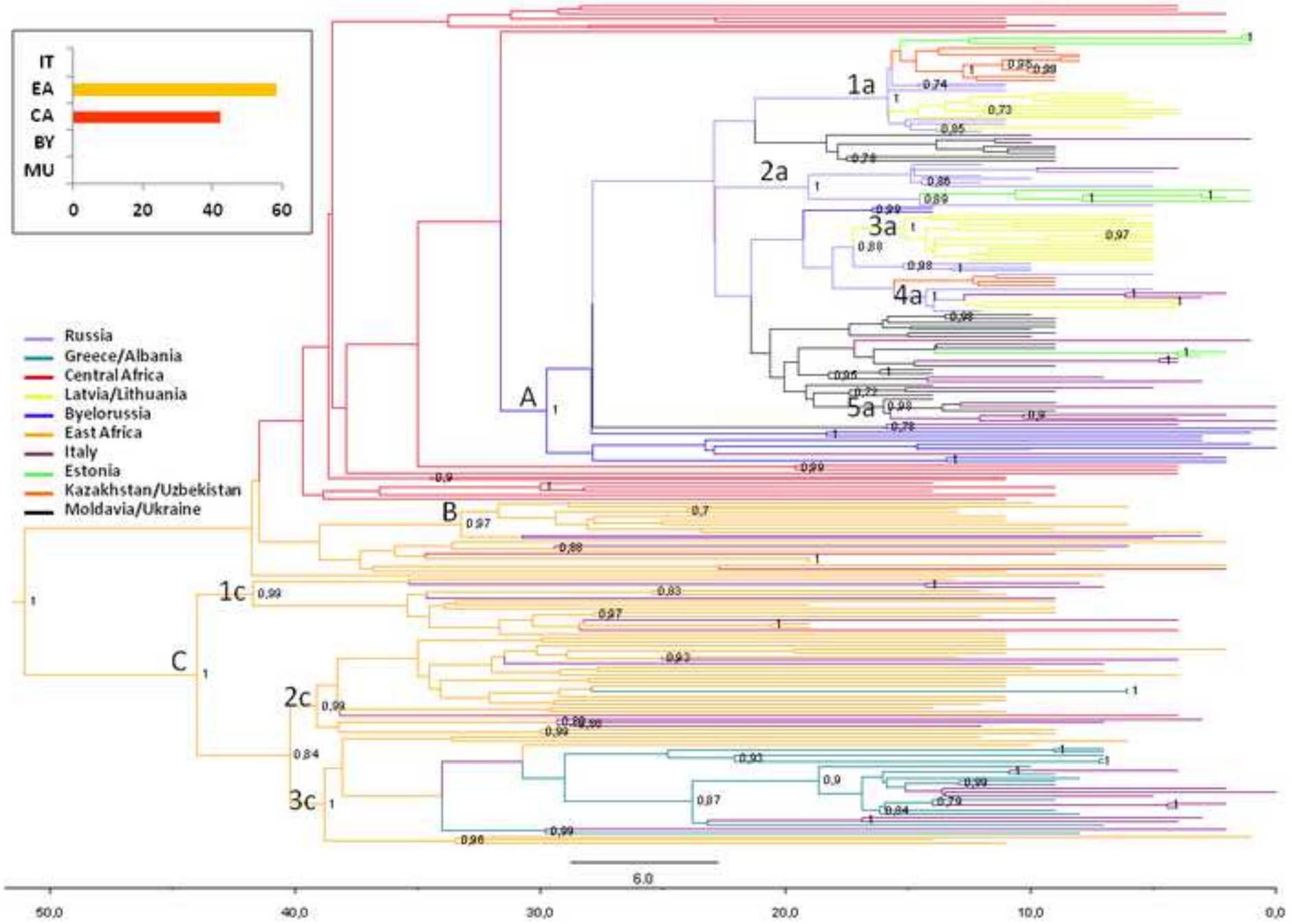
Figure

Click here to download Figure Figure 1.tif



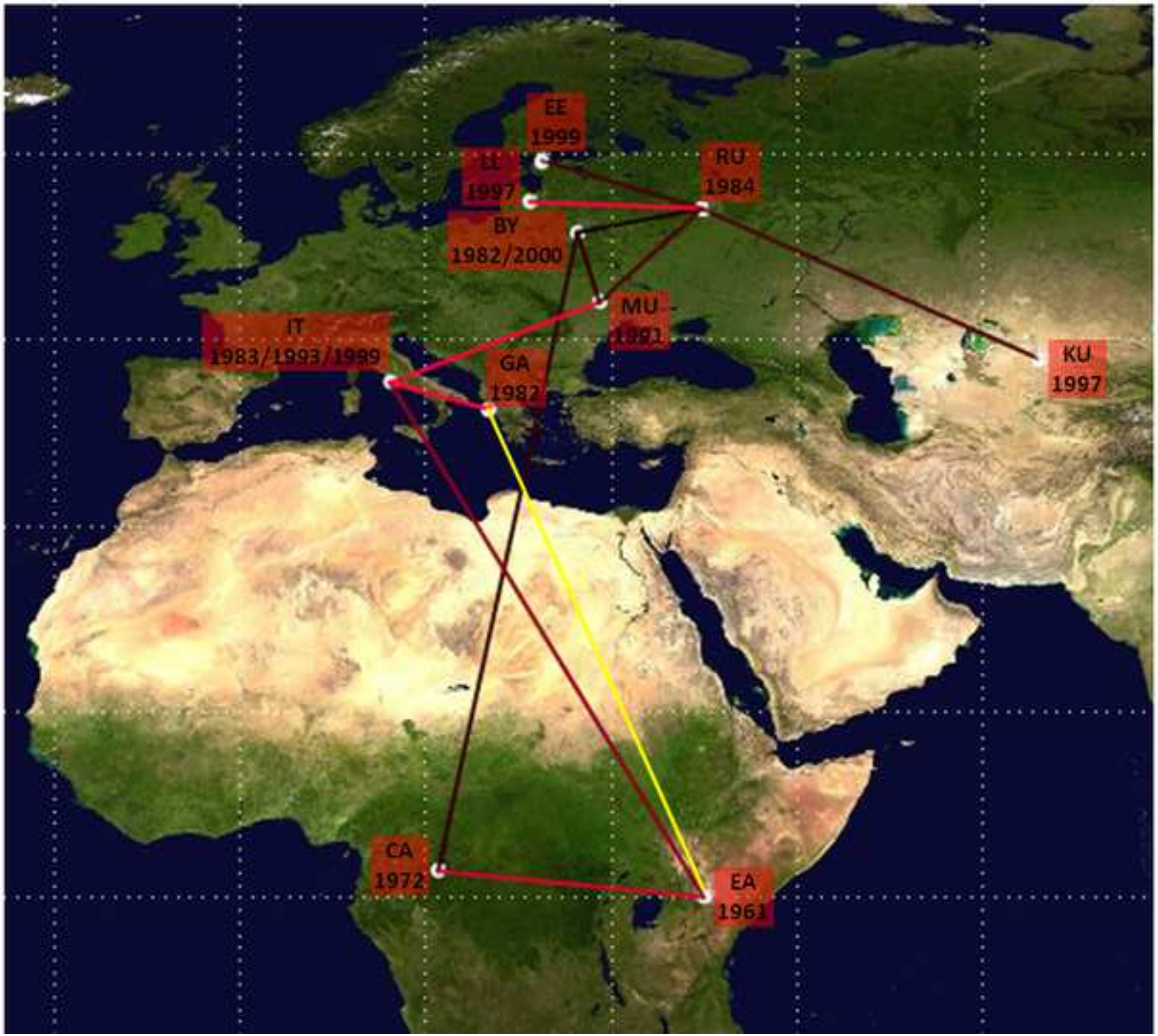
Figure

Click here to download Figure Figura 2.tif



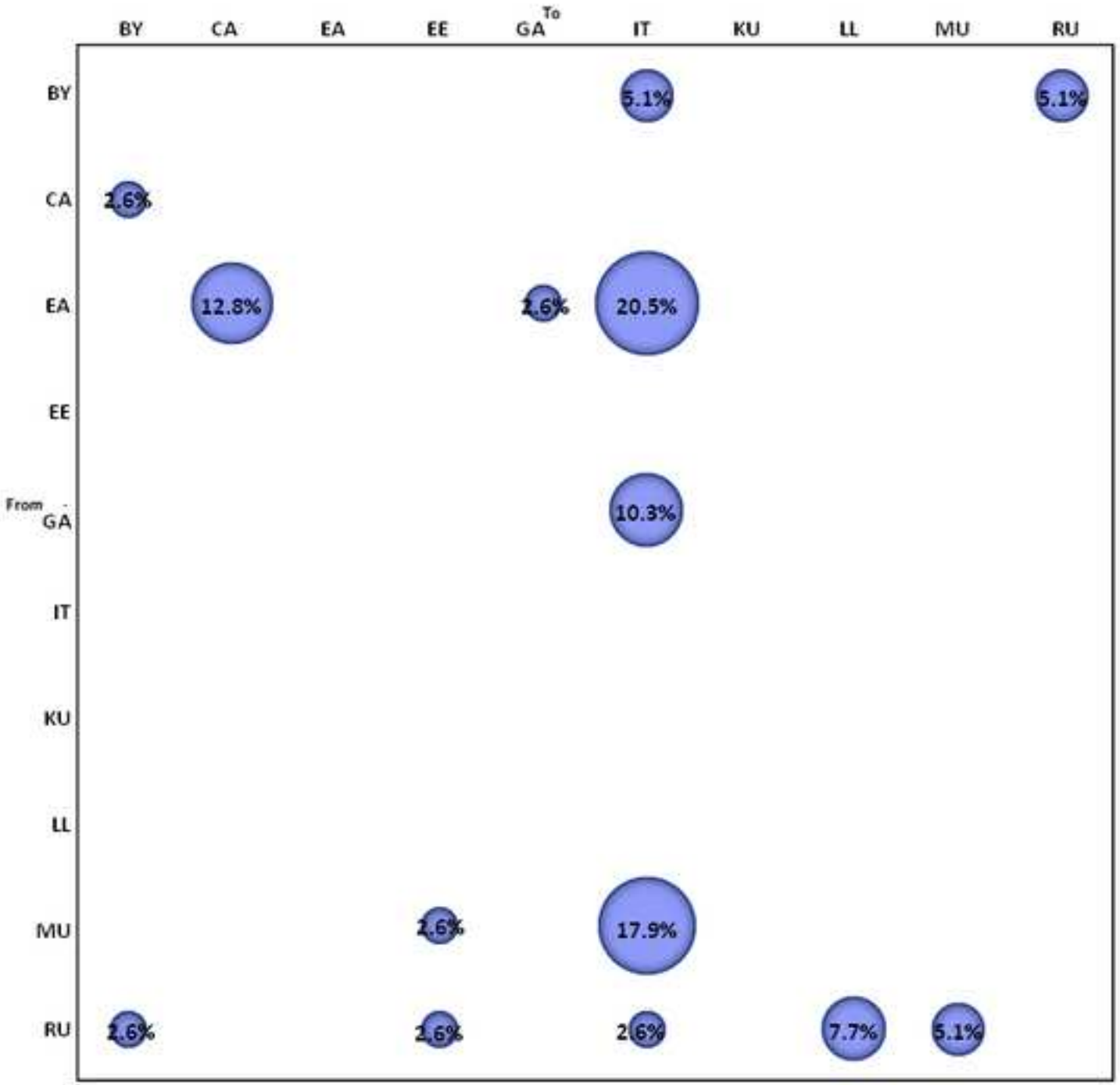
Figure

Click here to download Figure Figure 3.tif



Figure

Click here to download Figure Figure 4.tif



HIV-1 A1 subtype epidemic in Italy originated from Africa and Eastern Europe and shows a high frequency of transmission chains involving intravenous drug users

Alessia Lai[1]*, Giorgio Bozzi[1], Marco Franzetti[1], Francesca Binda[1], Francesco R Simonetti[1], Andrea De Luca[2], Valeria Micheli[3], Paola Meraviglia[4], Patrizia Bagnarelli[5], Antonio Di Biagio[6], Laura Monno[7], Francesco Saladini[8], Maurizio Zazzi[8], Gianguglielmo Zehender[1], Massimo Ciccozzi[9], Claudia Balotta[1]

[1]Department of Biomedical and Clinical Sciences "Luigi Sacco", Infectious Diseases and Immunopathology Section, 'L. Sacco' Hospital, University of Milan, Milan, Italy; [2]Division of Infectious Diseases, Siena University Hospital, Siena; [3]Laboratory of Microbiology, 'L. Sacco' Hospital, Milan, Italy; [4]2nd Division of Infectious Diseases, 'L. Sacco' Hospital, Milan, Italy; [5]Department of Biomedical Science, Section of Microbiology, Laboratory of Virology, University Politecnica delle Marche, Ancona, Italy; [6]Infectious Diseases Clinic, San Martino Hospital, Genoa; [7]Division of Infectious Disease, University of Bari, Bari, Italy; [8]Department of Medical Biotechnology, University of Siena, Siena, Italy; [9] Epidemiology Unit, Department of Infectious, Parasite and Immune-Mediated Diseases, Italian Institute of Health, Rome, Italy.

*Corresponding Author:

Dr. Alessia Lai, PhD: Department of Biomedical and Clinical Sciences "Luigi Sacco", Infectious Diseases and Immunopathology Section, University of Milan, Via G.B. Grassi 74, 20157 Milan, Italy – Phone (+39) 0250319775, Fax: (+39) 0250319768 – e-mail: alessia.lai@unimi.it

Abstract

Subtype A accounts for ~~about~~ only 12% of HIV-1 infections worldwide ~~while~~ but predominates in Russia and Former Soviet Union countries of Eastern Europe. After an early propagation via heterosexual contacts, this variant spread explosively among intravenous drug users. A distinct A1 variant predominates in Greece and ~~Albania which~~Albania, which penetrated directly from Africa. Clade A1 accounts for 12.5% of non-B subtypes in Italy, being the most frequent after F1 subtype. Aim of this study was to investigate the circulation of A1 subtype in Italy and trace its origin and diffusion through phylogenetic and phylodynamic approaches. The phylogenetic analysis of 113 A1 *pol* sequences included in the Italian ARCA database, indicated that 71 patients (62.8%) clustered within 5 clades. A higher probability to be detected in clusters was found for patients from Eastern Europe and Italy (88.9% and 60.4%, respectively) compared to those from Africa (20%) (*p<.001*). Higher proportions of clustering sequences were found in intravenous drug users with respect to heterosexuals (85.7% *vs.* 59.3%, *p=.056*) and in women with respect to men (81.4% ~~versus~~ *vs.* 53.2%, *p<.006*).

~~Dated~~ Subtype A1 dated phylogeny indicated an East African origin around 1961. Phylogeographical reconstruction highlighted 3 significant groups. One involved East European and some Italian variants, the second encompassed some Italian and African strains, the latter included the majority of viruses carried by African and Italian ~~strains~~ subjects and all viral sequences from Albania and Greece.

Subtype A1 originated in Central Africa and spread among East European countries in 1982. It entered Italy through three introduction events: directly from East Africa, from Albania and Greece, and from the area encompassing Moldavia and Ukraine. As in previously documented A1 epidemics of East European countries, HIV-1 A1 subtype spread in Italy in part through intravenous drug users. However, Eastern European women contributed to the penetration of such variant, probably through sex work.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Background

The evolution of HIV-1 subtype A, similar to subtype F, gave rise to different sub-subtypes (A1, A2, A3, A4) [1,2], defined as sub-clusters~~,~~ and formed by distinct lineages highly related to the ~~pertaining~~ parental clade. The genetic distance between these subclades is about half of that between subtypes.

Subtype A ~~sustains~~is responsible of about 80% of the HIV-1 epidemic in Former Soviet Union (FSU) countries [3-5]. ~~Evolved~~ Evolving from a Central African ancestor, this variant initially spread through heterosexual (HE) contact in Southern Ukraine. Subtype A then begun to spread dramatically among injective drug users (IDUs) in the port city of Odessa around 1995. During the following years, the HIV-1 epidemic sustained by subtype A among IDU (labeled A_{fsu} or *IDU-A*) propagated through other FSU countries such as Russia and Estonia [6]. The explosive spread of this clade within the population of FSU IDUs was due to effective transmission networks taking place in the favorable setting of the massive socio-economic crisis which followed the fall of the Soviet Union Republics [7]. In the largely populated city of St. Petersbourg the proportion of A_{fsu} infections increased from 37% in the 1997-2004 period to 91% in 2006 of HIV-1 infections [6].

A feature of the IDU-A variant is a high homogeneity of viral sequences, probably due to the high transmission rates among IDUs following single introduction events in distinct geographic areas and the recent spread of the epidemic ~~and high transmission rates among IDUs following single introduction events in distinct geographic areas~~ [8].

~~Following~~ After its explosive diffusion in Russia, subtype A spread in neighboring countries [9-12]. In Bulgaria, subtype A was isolated only in a single individual before 1995, however, it accounted for 27% infections ~~after the same~~ in further years [13], suggesting a late introduction through multiple events.

The prevalence of subtype A is approximately 2% in Western and Central Europe, however this variant has established extensive epidemics in some Mediterranean countries, such as Albania, Cyprus and Greece. In Greece, clade A1 is the most common non-B subtype (20.6%), rising from a 6% prevalence in 1984 to 42% in 2004 [14,15]. The introduction of this subtype in Greece dates back to the first epidemic phase in the

country (time of the most recent common ancestor, tMRCA, 1978), probably originating from Central Africa [16]. Differently from other European countries, subtype A is the most frequent in long-dated residents compared with subtype B, and was involved in sexual transmission risk groups more lately [14]. Similarly, the clade A epidemic in Albania probably arose from Greece, Albanian and Greek sequences being more related to African ones than to Eastern European ones [17].

Clade A is a parental subtype in most of ~~the~~ known circulating recombinant forms (CRFs), particularly in the most prevalent ones. These CRFs are estimated to sustain 27% HIV-1 infections globally [18], especially CRF02_AG in Western Africa [19], CRF01_AE in Thailand [20] and CRF03_AB in the FSU.

The high prevalence of co-circulating subtype A and B in Eastern Europe represented the background for the origin of CRF03_AB in Southern Ukraine, giving rise to an outbreak through IDUs in the city of Kaliningrad in 1996 [21].

Due to migration fluxes from Africa, FSU and South America, non-B subtype circulation is increasing in Italy as well as in all Western countries of Europe. The overall prevalence of infections ~~of~~ due to non-B clade ~~infections~~ in Italy ~~was recently reported as~~ is 11.4% having raised from 2.6% to 18.9% over three decades. Among these, subtype A is the second in prevalence (12.7%) after clade F1 (23.7%) [22]. The aim of this study was to investigate the features of A1 subtype circulation in Italy and trace its origin and diffusion through phylogenetic and phylodynamic approaches.

Formatted: English (United States)

Patients and methods

Study population

We studied 113 individuals carrying HIV-1 A1 subtype. Patients were sampled from 1999 through 2011. Subjects signed an informed consent to have their anonymized data stored on a central server of the ARCA database (www.dbarca.net) and used for research studies. Authors working in the clinical setting interacted with some of the patients included in the study as part of their own routine HIV care. No additional visits were scheduled, apart from those regularly planned for HIV monitoring, according to Italian national guidelines. ARCA is an observational HIV cohort approved by the Regional Ethical Committee of Tuscany (Comitato Etico Area Vasta Toscana Sudest). HIV-1 protease (PR) and partial reverse transcriptase (RT) sequences were generated by local centers for routine drug resistance testing at diagnosis or prior to the start of antiretroviral therapy or at treatment failure. Epidemiological data (gender, risk category, country of origin, date of diagnosis and age) were collected by physicians from patient medical records and then included in the databases together with virological, immunological and treatment information. Only the earliest available HIV-1 genotype was considered for each patient. The study was conducted in accordance with the 1964 Declaration of Helsinki and the ethical standards of the Italian Ministry of Health.

Phylogenetic dataset

The analysis of epidemiological clusters was performed on the starting dataset of 113 HIV-1 A1 subtypes.

The evolutionary population dynamics was estimated in the dataset of Italian patients (n=53).

A reference set of 229 subtype A1 *pol* gene sequences was obtained from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov). Viral sequences were selected according to the following inclusion criteria: i) sequences had already been published in peer-reviewed journals; ii) the subtype assignment of each sequences classified as non-recombinant was certain; iii) the city/state of origin and the year of sampling were known and clearly established in the original publication. The final reference set includes subtype A1 sequences of Central Africa (n=22; Democratic Republic of Congo=13, Central African Republic=3,

Gabon=6), East Africa (n=52; Uganda=17, Tanzania=16, Kenya=19), Estonia (n=9), Latvia (n=19), Lithuania (n=5), Moldavia (n=5), Ukraine (n=20), Byelorussia (n=10), Russia (n=20), Kazakhstan (n=6), Uzbekistan (n=7), Greece (n=9), Albania (n=8). The accession number list of phylogeographic references is provided in S1 Table ~~The accession number list of phylogeographic references is provided in S1 Table.~~ Thirty seven randomized sequences belonging to patients with established Italian citizenship present in a starting dataset were included. ~~The phylogeographic dataset of 229 isolates was used to evaluated the migration pattern of HIV-1 subtype A1 circulation.~~ The phylogeographic dataset of 229 isolates was used to evaluate the migration pattern of HIV-1 subtype A1 circulation.

The phylogenetic signal was evaluated on the phylogeographic dataset.

Alignment and phylodynamic analysis

The sequences were manually aligned using Bioedit software (v7.2.5; http://www.mbio.ncsu.edu/bioedit/bioedit.html). All sites of major antiretroviral drug resistance mutations in the protease and reverse transcriptase regions, identified through the last updated Surveillance Drug Resistance Mutations (SDRM) (http://hivdb.stanford.edu/cgi-bin/AgMutPrev.cgi) and International AIDS Society (IAS) (http://www.iasusa.org/guidelines/index.html) lists for naïve and treated patients respectively, were removed to avoid bias due to convergence related to drug resistance. The final alignment encompassed 834 nucleotides.

The Bayesian phylogenetic tree was reconstructed by means of both MrBayes and Beast using the general time reversible (GTR)+ gamma distribution (G)+ proportion of invariable sites (I) model, which was selected using the Modeltest ~~V~~v. 3.7 program. A Markov Chain Monte Carlo (MCMC) search was made for $10x10^6$ generations using tree sampling every 100th generation and a burn-in fraction of 50%. Statistical support for specific clades was obtained by calculating the posterior probability (pp) of each monophyletic clade, and a posterior consensus tree was generated after a 50% burn-in. Clades with a pp of 1 were considered epidemiologically related clusters.

Formatted: Not Highlight

Formatted: Not Highlight

Dated trees, evolutionary rates and population growth were co-estimated by using Beast v.1.8.0 [23] software (http://beast.bio.ed.ac.uk) using the previous model. Different coalescent priors (constant population size, exponential growth, logistic growth and Bayesian skyline plot) were tested using both strict and relaxed molecular clock models. Different clock models were compared by calculating the Bayes Factor (BF) [15]. The difference (in $\log_e$ space) of marginal likelihood between any two models is $\log_e$ BF. Evidence against the null model (i.e. the one with lower marginal likelihood) was indicated by: 6>(2 $\log_e$ BF)>2 (positive), 10>(2 $\log_e$ BF)>6 (strong) and (2 $\log_e$ BF)>10 (very strong) [16]. BF calculations were performed with Tracer v.1.5 software (http://beast.bio.ed.ac.uk/Tracer). The BF analysis showed that the relaxed lognormal molecular clock fitted the data better than the strict clock model (2lnBF=114.5). Chains were run for 50 million generations with sampling every 5,000 generations. The results were visualized in Tracer v.1.5 and uncertainty in the estimates was indicated by 95% highest probability density (HPD) intervals. Convergence was assessed based on the ESS (effective-sample size) value and only parameter estimates with ESS>300 were accepted. The maximum clade credibility (MCC) tree was then selected from the posterior tree distribution using TreeAnnotator v.1.8.0 available in the BEAST software package. Final trees were visualized and manipulated with FigTree v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Likelihood mapping analysis

In order to obtain an overall feature of the phylogenetic signal present in the HIV-1 A1 sequences, we made a likelihood-mapping analysis of 10,000 random quartets generated using TreePuzzle (Schmidt et al., 2002). A likelihood map consists of an equilateral triangle: each dot within the triangle represents the likelihoods of the three possible unrooted trees for a set of four sequences (quartets), randomly selected from the dataset. The dots close to the corners or at the sides respectively represent tree-like (fully resolved phylogenies in which one tree is clearly better than the others) or network-like phylogenetic signals (three regions for which it is not possible to decide between two topologies). The central area of the map represents a star-like signal (the region where the star tree is the optimal tree).

Phylogeographic analysis

Analyses were performed using Bayesian Skyline coalescent tree prior with a relaxed clock (uncorrelated Lognormal model). Chains were run for 100 million generations with sampling every 10,000. The results were visualized using Tracer v.1.5. The ESS value for each parameter was >300, indicating sufficient mixing of the Markov chain. Rates with a BF of >3 were considered well supported and formed the migration pathway. The MCC tree was selected from the posterior tree distribution after a 10% burn-in using the TreeAnnotator program, version 1.8.0. The final tree was visualized using FigTree version 1.4.0. To provide a spatial projection, the migration routes indicated by the tree were visualized using Google Earth (http://earth.google.com) and the SPREAD program (available at http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html).

Viral Gene Flow Analysis

The MacClade program version 4 (Sinauer Associates, Sunderland, MA) was used to test viral gene out/in using a modified version of the Slatkin and Maddison test [24]. A one-character data matrix was obtained from the original dataset by assigning a one-letter code, indicating the geographical origin, to each taxon in the tree. The putative origin of each ancestral sequence (i.e., internal node) in the tree was then inferred by finding the most parsimonious reconstruction (MPR) of the ancestral character using either the ACCTRAN or DELTRAN option. The final tree length, that is the number of observed viral gene flow events in the genealogy, is computed and compared to the tree-length distribution of 10,000 trees obtained by random joining-splitting (null distribution). Observed genealogies significantly shorter than random trees indicate the presence of subdivided populations with restricted gene flow. The viral gene flow (migrations) was traced with the 'State changes and stasis tool' (MacClade software), which counts the number of changes in a tree for each pairwise character state. Gene flow was also calculated for the null distribution to assess

whether the gene flow events observed in the actual tree were significantly higher (>95%) or lower (<95%) than the values in the null distribution at the $p = 0.05$ level.

Formatted: Font: Italic

Statistical methods

Categorical variables were summarized as frequencies and comparisons between groups were performed using the χ2 or the Fisher exact test. For all analyses an α error of 5% was considered. The crude and Mantel–Haenszel-adjusted odds ratios of inclusion in clusters with 95% CIs were calculated in univariate and multivariate analyses. Analyses were performed using the SPSS software package (v.21.0, SPSS Inc. Chicago, IL, USA).

Results

Patient characteristics

One-hundred thirteen HIV-1 A1 infected patients with sequence data spanning a 13-year period (Table 1) were included. Italians, East Europeans, Africans, South Americans and Asians accounted for 46.9% (n=53), 15.9% (n=18), 8.8% (n=10), 1.8% (n=2) and 0.1% (n=1) of patients with known country of origin, respectively. Known risk factors were distributed as follows: 64.3% (n=27) HEs, 19% (n=8) men having sex with men and 16.7% (n=7) IDUs. Males accounted for 59% of patients (n=62). Median age was 35 years (range: 24-44 years). Naïve patients accounted for 21.2% (n=24); only two patients harbored resistant strains (8.3%) and only mutations selected by Protease Inhibitors (PI) were present.

Associations between subtype A and patients characteristics

~~The phylogenetic trees were constructed on 113 individuals carrying HIV-1 A1 subtype.~~ The phylogenetic trees were constructed on 113 individuals carrying HIV-1 A1 subtype. No differences were observed in the topology of trees obtained using MrBayes or BEAST programs. The BEAST tree of HIV-1 A1 sequences is shown in S1 Fig.

~~Based on this tree we identified 5 highly supported clades~~ Based on the latter tree we identified 5 highly supported clades (with posterior probability, pp, equal to 1) (S1 Fig) that included 71 patients (62.8%). A ~~single~~ large network (#5) encompassed 59 isolates (83.1% of clustering sequences), while the remaining four included three patients each. The major clade involved all patients from Eastern Europe and included 47.2% of Italians (n=25). Noteworthy, ~~In cluster #5~~ female gender accounted for 54.2% (n=32) in cluster #5. ~~Of note~~Differently to the other networks, cluster #4 involved only Italian subjects. Clusters #1 and #4 included only treated subjects. None of the clusters ~~was homogenous for~~included exclusively ~~subjects~~subjects with the same gender.

The overall epidemiological characteristics and those of patients included or not within epidemiological networks are described in Table 1.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Patients from Eastern Europe and Italy had a higher probability to be detected in clusters (n=16, 88.9% and n=32, 60.4%, respectively) compared to those from Africa (n=2, 20%) (*p<.001*). The proportion of clustering sequences was slightly higher in IDUs (n=6, 85.7%) and HEs (n=16, 59.3%) compared to MSM (n=2, 25%) (*p=.056*). Women carried a higher proportion of clustering sequences (n=35, 81.4%) compared to men (n=33, 53.2%) (*p<.006*). Among clustering females with known risk factor, 80% (n=8) were HEs and 20% IDUs. Seven of HE females were from East Europe (87.5%).

Male gender was predominant both in total population and in Italian patients within clusters (n=38, 71.7% *vs.* n=19, 59.4%, respectively) (*p<.001*), whereas female gender was prevalent among patients from Eastern Europe (n=13, 16.7% *vs.* n=15, 18.7%) (*p=.029*). No difference was observed in the distribution of naïve subjects inside or outside clusters, however 19 (79%) naïve individuals were present in the epidemiological chains. Patients involved in such chains were significantly younger (33 years, range: 28.5-40 years) than patients ~~laying~~ outside ~~the~~ clusters (39 years, range: 34-53.5 years) *(p=.003)*.

Phylodynamic analysis

~~The phylodynamic analysis was conducted on the dataset of patients with certain Italian citizenship (n=53) patients carrying HIV-1 A1 subtype.~~ The phylodynamic analysis was conducted on the dataset of patients with certain Italian citizenship (n=53) carrying HIV-1 A1 subtype.

The Bayesian Skyline plot showed the changes in population size at different times from the root of the tree to the time of the most recent isolates (year 2011). The mean substitution rate for the analyzed dataset was 1.11×10^{-3} (95% HPD: 2.82×10^{-4} – 1.87×10^{-3}). The effective number of infections grew exponentially approximately between 1990 and 1998 and reached a plateau around 2000, when the effective number of infections remained constant until the last calendar years (Fig 1).

Phylogeographical reconstruction

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Not Highlight

Formatted: Not Highlight

The likelihood mapping analysis of the HIV1-A1 dataset of 229 isolates showed that the percentage of dots falling in the central area of the triangles was 4.2%, thus indicating a fully resolved phylogenetic signal (S2 Fig).

The phylogeographic mMaximum clade credibility (MCC) tree, obtained grouping the 229 isolates within ten discrete groups on the basis of the sampling location (IT, Italy; CA, Central Africa; EA, East Africa; EE, Estonia; LL, Latvia/Lithuania; MU, Moldavia/Ukraine; BY, Byelorussia; RU, Russia; KU, Kazakhstan/Uzbekistan; GA, Greece/Albania) is shown in Fig 2.

The most probable location of the root of the tree was East Africa, supported by a state posterior probability (spp) significantly higher than that of the other locations (spp=0.58 *vs.* spp=0.42 for Central Africa, the second most probable location of the root). The date of the most recent common ancestor (MRCA), corresponding to the East African root, was estimated to be 51 years (95% HPD=42.1-69.4 years), corresponding to the year 1960.

Phylogeographical reconstruction highlighted ~~three~~ 3 highly supported groups (A, pp=1; B, pp=0.97; and C, pp=1). Group A involved all sequences from Eastern Europe and the most probable location is Byelorussia (spp=0.76). Sixteen Italian sequences were interspersed within this group. The mean tMRCA of the A group was 29.7 years, (95% HPD ~~range~~= 22-35.7 years) corresponding to 1981. In this group five clusters can be identified (1a, pp=1; 2a, pp=1; 3a, pp=1; 4a, pp=1; and 5a, pp=0.98). The first cluster (1a) included 27 isolates~~,~~ (10 from Kazakhstan/Uzbekistan, 3 from Estonia, 6 from Russia and 8 from Latvia/Lithuania); cluster 2a involved 12 strains~~,~~ (7 from Russia, 4 from Estonia and 1 from Italy); cluster 3a contained ~~13~~ only sequences from Latvia/Lithuania (n=13); cluster 4a contained 7 isolates (~~,~~ 3 from Latvia/Lithuania, 2 from Russia and 2 from Italy); cluster 5a included 7 isolates (~~,~~ 4 from Italy and 3 from Moldavia/Ukraine). The mean tMRCA of these clusters range from 15-19 years (95% HPD ~~range~~= 12-20 years).

Group~~s~~ B involved 9 African sequences ~~(spp=0.97 for East Africa)~~ and two Italian strains and the most probable location is East Africa (spp=1). The mean tMRCA of B group was 41.8 years (95% HPD ~~range~~= 32.8-42.7 years).

Group C involved the majority of African sequences (spp=1 for East Africa), 18 Italian sequences, all strains from Greece/Albania and 2 Central African sequences. The mean tMRCA of group C was 44 years, (95% HPD ~~range=~~ 36.1-47.8 years) corresponding to 1967. In this group three clusters were identified (1c, pp=0.99; 2c, pp=0.99; and 3c, pp=1). The first cluster (1c) included 16 isolates~~,~~ (11 from East Africa, 4 from Italy and 1 from Central Africa); the cluster 2c involved 27 strains~~,~~ (19 from East Africa, 5 from Italy and 1 from Central Africa); cluster ~~2c~~ 3c contained 30 sequences (~~,~~ 6 from East Africa, 15 from Greece/Albania and 9 from Italy). The mean tMRCA of these clusters range from 38-42 years (95% HPD ~~range~~= 32.2-44 years). tMRCA values, confidence intervals and the locations of the main groups and clusters are listed in Table 2.

Significant Linkages~~, Migration Rates and Fluxes~~

Bayesian phylogeographic analysis ~~s~~of 229 ~~hiv~~HIV-1 A1 isolates showed eleven non-zero rates between ten discrete locations (Italy, East Africa, Central Africa, Estonia, Latvia/Lithuania, Moldavia/Ukraine, Byelorussia, Russia, Kazakhstan/Uzbekistan and Greece/Albania). Significant linkage (BF>3.0) was found between Byelorussia and Moldavia/Ukraine (BF=6), Central Africa (BF=7.1) and Russia (BF=3.2). Russia was significantly linked with Estonia (BF=17), Kazakhstan/Uzbekistan (BF=20), Latvia/Lithuania (BF=32778.4) and~~,~~ Moldavia/Ukraine (BF=26.8). Italy was significantly linked with three locations: East Africa (BF=297.1), Greece/Albania and Moldavia/Ukraine (BF=32778.4 for both locations). Central Africa and East Africa are linked with a BF=5660.

The pathways of HIV-1 A1 dispersion are showed in Fig 3. HIV-1 A1 subtype originated in East Africa around 1961 and spread to Central Africa in 1972. Subsequently, the virus spread in East Europe starting from Byelorussia where the virus was introduced in 1982 from Central Africa. The virus spread to the Kazakhstan/Uzbekistan from Russia around 1996 ~~from Russia~~.

The HIV-1 A1 subtype reached Italy primarily from East Africa (year 1983) and subsequently from Greece/Albania and Moldavia/Ukraine around 1993-1999. A probable link related East Africa and

Formatted: Font: Calibri

Greece/Albania where the virus was introduced in 1983, even though a not significant BF could be observed.

Migration Rates and Fluxes

~~The migration pattern of HIV-1 subtype A1 circulation was calculated using the phylogeographic dataset of 229 isolates.~~ The migration pattern of HIV-1 subtype A1 circulation was calculated using the phylogeographic dataset of 229 isolates. The gene flow (migration) between different geographic areas was investigated with a modified version of the Slatkin and Maddison method. The null hypothesis of panmixia (i.e. no population subdivision or complete intermixing of sequences from different geographical areas) was tested using a bubblegram. The migration flow among the five locations was then estimated as observed migration in the genealogy (Fig 4). The null hypothesis of panmixia was ejected by the randomization test ($p = ~~0~~.0001$) for all the countries in the bubblegram. ACCTRAN resolving option~~s~~ was used for uncertainties present in the resulting tree.

HIV-1 A1 gene flows were observed from East Africa to Central Africa (12.8%), Greece/Albania (2.6%) and Italy (20.5%). Italy received migrations also from Byelorussia (5.1%), Greece/Albania (10.3%), Moldavia/Ukraine (17.9%) and Russia (2.6%).

The gene flow from Central Africa to Byelorussia, from Byelorussia to Russia and from Moldavia/Ukraine to Estonia accounted for 2.6%, 5.1% and 2.6% of the observed migrations, respectively (Fig 4). HIV-1 subtype A1 gene flow ~~highlighted the role of~~indicated Russia as the epicenter of HIV-1 A1 epidemic in East Europe, branching out to Byelorussia (2.6%), Moldavia/Ukraine (5.1%) and Latvia/Lithuania (7.7%) (Fig 4).

Formatted: Not Highlight

Formatted: Font: Italic

~~Discussion~~ Discussion

The prevalence of non-B subtypes has increased over time in Italy as well as in Europe and this rise can be attributed mainly to an increasing frequency of F1, C, CRF02_AG and A1 subtypes [22].

The present study is the first detailed investigation of the HIV-1 A1 molecular epidemiology in Italy using a national database and a Bayesian phylogenetic approach.

An extraordinary proportion of variants were associated to A1 epidemiological networks (more than 62% of studied sequences). We observed a relevant role of heterosexual contacts and injecting drug use. Italian male heterosexuals and IDUs ~~males~~ predominated while foreign-born patients were mainly heterosexual females from East Europe. The high frequency of Italian heterosexual males and Eastern Europe females within epidemiological networks indicated that sexual ~~intercourses~~ contacts have led to a relevant circulation of subtype A1 in Italy. Only one cluster involving patients with Italian citizenship was identified, while the majority of Italian strains were intermixed within East European patients. Additionally, despite the reported decrease of infections due to injecting drug usage (http://www.iss.it/binary/ccoa/cont/ONLINE_6_COA_2014.pdf), the circulation of A1 subtype partially remains related to this modality of infection as in previously documented A1 epidemics of East European countries [21]. In contrast with previous published data regarding other non-B subtypes circulating in Italy, such as subtype C and F1 which mobility is associated with ~~man~~ men having sex with men, the A1 subtype remains related to heterosexuals and IDUs [25,26].

The present study also investigated the HIV-1 A1 epidemics in Italy and worldwide, with the aim to establish the origin of the A1 variant circulating in Italy as well as in the European and African context. The present phylogenetic analysis showed that subtype A1 ~~probably~~ originated, as expected, in East Africa (where subtype A is the most prevalent clade) with a recent epidemic, originating around 1961 [27].

Global subtype A1 dispersion showed a strong geographic compartmentalization in many clusters. Dispersal routes of subtype A1 from Africa to Europe were probably associated with immigration from Africa while viral mobility within Eastern Europe is probably due to ~~IDU~~ connections among IDUs. After its origin in East Africa subtype A1 migrated in Central Africa; both variants from East and Central Africa likely originated the epidemic in Europe. The spread among East European countries started from 1982, corresponding to the identification of the first outbreaks in countries formerly part of Soviet Union as reported by literature [7].

In general, several subtype A migration pathways exist among European countries which are mainly associated with heterosexuals and IDUs in Eastern Europe. Greece and Albania are a consistent exception due to the establishment of a large HIV-1 subtype A local epidemic sustaining 20.6% and 56.1% of infections, respectively [13,14]. In agreement with data of Paraskevis et al. [14], our findings showed a monophyletic origin of subtype A epidemic in Greece and Albania and confirmed its origin in East Africa (Kenya and Tanzania) with a date of origin around 1983.

Phylogeographical reconstruction suggests that A1 clade entered Italy through three distinct introduction events: an early first getaway took place directly from East Africa, where subtype A has a high prevalence, while two ~~more~~ additional events occurred from the Southern Balkan Peninsula and the area encompassing Moldavia and Ukraine. This phenomenon could be related to Eastern European women ~~that~~ who contributed to the penetration of such variant probably through sex work~~ers~~ as demonstrated by the high proportion of ~~HE females~~ women in clusters. It has been documented that, during the past 10 years, Albanian female sex workers traveled to other Balkan countries and to Western Europe, especially Italy and Greece. The very low percentage of African subjects within epidemiological networks suggests a minor role of the African A1 variant compared to the East European one in the circulation of this subtype in Italy. ~~Moreover, T~~this is in agreement with the high degree of homogeneity of the virus populations carried by IDUs in the former Soviet Union countries which suggests that they shares a quite recent common ancestor, as a result of a single introduction. In contrast, HIV-1 A1 African sequences presented a high

genetic variability, that increased in Democratic Republic of Congo in comparison to in ~~w~~Western~~ern~~ or Eeastern~~ern~~ Africa [11,16,28].

The estimation of the demographic history of subtype A1 in Italy revealed that the epidemic growth during the 1990s. The stabilization of the epidemic around 2000 can be probably explained by the reduced number of infections related to injecting drug usage and the diffusion of combination antiretroviral therapy in infected IDUs.

Globally, we ~~could~~ hypothesize three different coeval introduction events of the A1 variant from Africa into Europe; the former involving Byelorussia [29], the second Greece and the latter Italy. The migration fluxes confirmed the results of phylogeographic reconstruction giving a strong support to the link between East Africa and Greece.

Some possible limits and biases of the present study ~~should~~ have to be mentioned. First, although our dataset includes sequences from a national database we cannot exclude the possibility that the study population is not truly representative of all HIV-1 A1 subtype cases in Italy. Another ~~possible~~ potential limit of our study is that we analyzed a portion of *pol* gene, rather than full genome sequences, which could underestimate recombinant strains that frequently include A1 subtype portions (such as CRF01_AE or CRF02_AG, very frequent in Italy). Nevertheless, phylogenetic and recombination analysis were performed to confirm the subtype assignment on sequences classified as A1 subtype in the ARCA database by also comparing these strains with ~~a~~other known CRFs. Additionally, the relative lack of country of origin and risk factor information for some patients could have weakened the strength of the observed associations.

In conclusion, this study provides novel information about the A1 subtype epidemic and its geographic origin, as well as the routes of transmission among patients living in Italy and carrying such variant. Due to its founder effect, subtype B predominates the HIV-1 Italian epidemic. However, the geographic position of Italy in the center of Mediterranean basin and the increase of population mobility in Europe, had the potential to modify the previous HIV-1 epidemiological structure. The findings in the present study

389 contribute to the understanding of this changing pattern ad highlight the importance of an extensive and
390 continuous monitoring of the spread of HIV-1 variants in Europe.
391

FIGURE LEGENDS

Fig 1. Skyline plot obtained by analyzing the data set of Italian patients (n=53). Ordinate: the number of effective infections at time t (Ne(t)); abscissa: time (in years before the present). The thick solid line represents the median value and the grey area the 95% HPD of the Ne(t) estimates. The vertical lines indicate the 95% lower HPD (dotted) and the mean tMRCA estimate (bold) of the tree root.

Fig 2. Bayesian phylogeographical tree of 229 HIV-1 A *pol* sequences with branches colored on the basis of the most probable location of the descendent nodes. The correspondences between the locations and colors are shown in the panel (left), and the posterior probabilities >0.8 are indicated on the internal nodes of the tree. The scale axis below the tree shows the number of years before the last sampling time.

Fig 3. Significant non-zero HIV-1 A1 migration rates worldwide. Only the rates supported by a BF of >3 are shown in red, yellow indicates the probable rate. The relative strength of the statistical support is indicated by the color of the lines (from dark red , *id est* weak to light red *id est* strong). The map was reconstructed using SPREAD program. This figure is similar but not identical to the original image, and is therefore for illustrative purposes only.

Formatted: Font: Italic

Formatted: Font: Italic

Fig 4. Migration pattern of HIV-1 subtype A1 circulation based on phylogeographic dataset. The bubblegram shows the frequency of gene flow (migrations) to/from different geographic areas. The surface of each circle is proportional to the percentage of observed migrations in the Maximum Likelihood genealogy. Migrations were inferred with a modified version of the Slatkin and Maddison algorithm. BY, Byelorussia; CA, Central Africa; EA, East Africa; EE, Estonia; GA, Greece/Albania; IT, Italy; KU, Kazakhstan/Uzbekistan; LL, Latvia/Lithuania; MU, Moldavia/Ukraine; RU, Russia.

417 Table 1. Characteristics of 113 patients carrying subtype A1

418

		All patients	Patients within clusters	Patient outside clusters	*p*
ORIGIN % (n)	**Italy**	46.9 (53)	60.4 (32)	39.6 (21)	
	East Europe	15.9 (18)	88.9 (16)	11.1 (2)	
	Africa	8.8 (10)	20.0 (2)	80.0 (8)	*p<.001*
	South America	1.8 (2)	100 (2)	0.0 (0)	
	Asia	0.1 (1)	100 (1)	0.0 (0)	
	Unknown	26.5 (29)	96.6 (28)	3,4 (1)	
RISK FACTOR % (n)	**HEs**	64,3 (27)	59.3 (16)	40.7 (11)	
	MSM	19.0 (8)	25.0 (2)	75.0 (6)	*p=.056*
	IDUs	16,7 (7)	85.7 (6)	14.3 (1)	
GENDER % (n)	**Male**	59.0 (62)	53.2 (33)	46.8 (29)	*p<.006*
	Female	38.2 (97)	81.4 (35)	18.6 (62)	
AGE (median years)		35	33	39	

419

420

421 Table 2. tMRCAs, credibility intervals (95%HPD) and location of the main groups (A, B, C) and clades (1a, 2a,
422 3a, 4a, 5a, 1c, 2c, 3c) on the basis of the dataset of 229 HIV-1 subtype A1 isolates.

423

	tMRCA	95% HPD	Location
Root	51	42.1-69.4	East Africa
A	30	21.9-35.7	Byelorussia
1a	16	15.4-20	Russia
2a	19	14.2-18.4	Russia
3a	15	13.8-18.6	Latvia/Lithuania
4a	16	dic-17	Russia
5a	16	14.4-18.3	Moldavia/Ukraine
B	42	32.8-42.7	East Africa
C	44	36.1-47.8	Central Africa
1c	42	32.8-42.7	East Africa
2c	39	34-44	East Africa
3c	39	32.2-42.4	East Africa

424

425 tMRCA: time of the most recent common ancestor; HPD: highest posterior density.

426

ACKNOWLEDGMENTS

The authors thank all patients included in the study.

SUPPORTING INFORMATIONS

S1 Table. Accession numbers of used references.

S1 Fig. Bayesian tree of 113 HIV-1 A1 isolates obtained with Beast program. The 5 supported clusters are highlighted in red color. Node labels indicate posterior probability values (pp). Only values higher than 0.8 are shown.

S2 Fig. Likelihood mapping of the HIV-1 A1 data sets of 229 nucleotide sequences obtain using TreePuzzle program. The three corners represent fully resolved tree topologies, id est the presence of a tree-like phylogenetic signal in the given data set.

References

1. Gao F, Vidal N, Li Y, Trask SA, Chen Y, et al. (2001) Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res Hum Retroviruses 17: 675-688.

2. Meloni ST, Kim B, Sankale JL, Hamel DJ, Tovanabutra S, et al. (2004) Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. J Virol 78: 12438-12445.

3. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25: 679-689.

4. Galkin AN, Bychenko AB, Kuznecovs G, Filinova EY (2012) Isolation, characterization, and sequence analysis of a novel HIV type 1 subtype A1 strain from Russia. AIDS Res Hum Retroviruses 28: 1139-1146.

5. Smolskaya T, Liitsola K, Zetterberg V, Golovanova E, Kevlova N, et al. (2006) HIV epidemiology in the Northwestern Federal District of Russia: dominance of HIV type 1 subtype A. AIDS Res Hum Retroviruses 22: 1074-1080.

6. Thomson MM, Vinogradova A, Delgado E, Rakhmanova A, Yakovlev A, et al. (2009) 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 51: 332-339.

7. Thomson MM, de Parga EV, Vinogradova A, Sierra M, Yakovlev A, et al. (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.

8. Lukashov VV, Huismans R, Rakhmanova AG, Lisitsina ZN, Akhtyrskaya NA, et al. (1999) Circulation of subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St. Petersburg, Russia, correlates with geographical origin of infections. AIDS Res Hum Retroviruses 15: 1577-1583.

Formatted: Line spacing: Double

Formatted: Italian (Italy)

Formatted: Italian (Italy)

9. Dehne KL, Pokrovskiy V, Kobyshcha Y, Schwartlander B (2000) Update on the epidemics of HIV and other sexually transmitted infections in the newly independent states of the former Soviet Union. Aids 14 Suppl 3: S75-84.

10. Laga V, Vasilyev A, Lapovok I, Grigoryan S, Papoyan A, et al. (2015) HIV Type 1 Subtype A1 Dominates in Armenia. Curr HIV Res 13: 219-225.

11. Pandrea I, Descamps D, Collin G, Robertson DL, Damond F, et al. (2001) HIV type 1 genetic diversity and genotypic drug susceptibility in the Republic of Moldova. AIDS Res Hum Retroviruses 17: 1297-1304.

12. Grossman Z, Avidor B, Mor Z, Chowers M, Levy I, et al. (2015) A Population-Structured HIV Epidemic in Israel: Roles of Risk and Ethnicity. PLoS One 10: e0135061.

13. Salemi M, Goodenow MM, Montieri S, de Oliveira T, Santoro MM, et al. (2008) The HIV type 1 epidemic in Bulgaria involves multiple subtypes and is sustained by continuous viral inflow from West and East European countries. AIDS Res Hum Retroviruses 24: 771-779.

14. Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Paparizos V, et al. (2007) Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. J Infect Dis 196: 1167-1176.

15. Antoniadou ZA, Kousiappa I, Skoura L, Pilalas D, Metallidis S, et al. (2014) Short communication: molecular epidemiology of HIV type 1 infection in northern Greece (2009-2010): evidence of a transmission cluster of HIV type 1 subtype A1 drug-resistant strains among men who have sex with men. AIDS Res Hum Retroviruses 30: 225-232.

16. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, et al. (2000) Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J Virol 74: 10498-10507.

17. Ciccozzi M, Gori C, Boros S, Ruiz-Alvarez MJ, Harxhi A, et al. (2005) Molecular diversity of HIV in Albania. J Infect Dis 192: 475-479.

Formatted: Italian (Italy)

18. Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J (2002) Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J Acquir Immune Defic Syndr 29: 184-190.

19. Andersson S, Norrgren H, Dias F, Biberfeld G, Albert J (1999) Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from guinea-bissau with single or dual infections: predominance of a distinct HIV-1 subtype A/G recombinant in West Africa. Virology 262: 312-320.

20. Phuphuakrat A, Pasomsub E, Kiertiburanakul S, Chantratita W, Sungkanuparph S (2012) HIV type 1 integrase polymorphisms in treatment-naive and treatment-experienced HIV type 1-infected patients in Thailand where HIV type 1 subtype A/E predominates. AIDS Res Hum Retroviruses 28: 937-943.

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

22. Lai A, Riva C, Marconi A, Balestrieri M, Razzolini F, et al. (2010) Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV Med 11: 593-602.

23. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

24. Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123: 603-613.

25. Lai A, Bozzi G, Franzetti M, Binda F, Simonetti FR, et al. (2014) Phylogenetic analysis provides evidence of interactions between Italian heterosexual and South American homosexual males as the main source of national HIV-1 subtype C epidemics. J Med Virol 86: 729-736.

Formatted: Italian (Italy)

Formatted: Italian (Italy)

Formatted: Italian (Italy)

26. Lai A, Simonetti FR, Zehender G, De Luca A, Micheli V, et al. (2012) HIV-1 subtype F1 epidemiological networks among Italian heterosexual males are associated with introduction events from South America. PLoS One 7: e42223.

27. Njai HF, Ewings FM, Lyimo E, Foulongne V, Ngerageza D, et al. (2013) Deciphering the complex distribution of human immunodeficiency virus type 1 subtypes among different cohorts in Northern Tanzania. PLoS One 8: e81848.

28. Nyamache AK, Muigai AW, Khamadi SA (2013) Circulating trends of non-B HIV type 1 subtypes among Kenyan individuals. AIDS Res Hum Retroviruses 29: 400-403.

29. Nabatov AA, Kravchenko ON, Lyulchuk MG, Shcherbinskaya AM, Lukashov VV (2002) 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 18: 891-895.

535
536
537
538
539
540

Formatted: Justified

Reviewer #1

The manuscript was revised to improve English. Regarding the difficulty in understanding the manuscript it has to be taken into account that phylogenetic studies are complex and sophisticated and need to be conducted with ad hoc biostatistical and mathematical methods. We tried when possible to clarify the performed analysis, while the usage of technical language must be applied. We, as well as other research groups, published few papers in this field on Plos One using the same language and depth of details of our experimental procedures (Zehender G, et al. PLoS One. 2012; Lai A, et al. PLoS One. 2012; Zehender G, et al. PLoS One. 2012; Zehender G, et al. PLoS One. 2013; Frentz D, et al. PLoS One. 2014).

Specific comments:

1) Second paragraph (line 200 on) – no indication of what is analyzed in this paragraph. S1 figure totally unclear, including the legend.

The introduction of the second paragraph, as well as the legend of S1 figure, were modified to indicate what is analyzed.

2) Phylodynamic analysis – no indication of what is analyzed in this paragraph. Figure 1 and legend are totally unclear.

The introduction of the this paragraph was modified.

3) Phyleogeographical reconstruction – what is S2 figure? What should the reader understand? The tree in Fig. 2 is from 229 isolates. Should Authors explain that these isolates are from the Los Alamos Database? Do they include the ones analyzed in the present study? What is a "MCC tree"?

As described in Patients and methods, regarding the Phylogenetic dataset (line 102-115), the phylogenetic signal was evaluated on phylogeographic dataset of 229 subtype A1 pol gene sequences obtained either from the Los Alamos HIV Sequence Database (n=119) (www.hiv.lanl.gov) or from Italian patients (n=37). S2 Figure represent the result of TreePuzzle program as described in Patients and methods- Likelihood mapping analysis (line 147-155). Additional informations have been added in the figure legend. The MCC tree is the maximum clade credibility tree as indicated at line 143.

4) Figure 3 – is this coming from the literature? Did the Authors generate it? If yes, how. This is totally unclear.

We generated Figure 3 using SPREAD program, as reported in Patients and methods: 'To provide a spatial projection, the migration routes indicated by the tree were visualized using Google Earth (http://earth.google.com) and the SPREAD program (available at http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html)'.
The original Figure 3 was removed and replaced with a new one.

5) Last paragraph and Figure 4 – no explanation whatsoever.

More details have been added on legend of Figure 4 and in the Patients and methods (Phylogenetic dataset paragraph).

Reviewer #2: In this manuscript Lai et al are using a relevant number of samples to perform an in-depth analysis of the HIV-1 subtype A strain diversity in Italy. They report that, similar to other Western countries in which epidemic spread is governed by founder effect, >60% of the subtype A strains from Italy cluster in 5 subclades that can be traced to Eastern Europe. They also report that, in agreement with previously published data, sequences from European patients have a higher probability to be grouped in specific clusters than those from African patients and that a higher proportion of clustering sequences can be found in intravenous drug users than in heterosexuals and in women with respect to men. These results were expected and are in agreement with known features of molecular epidemiology of HIV-1. They also dated the origin of Subtype A in Italy and established that the strains most likely emerged from East Africa around 1961.

Altogether, the paper is pretty straight forward and of some interest for the field of HIV diversity. Data might be built into the overall global picture of HIV diversity.

We thank the reviewer for the comments and the precise revision that improve the quality of the present version of the manuscript. We addressed specific issues as follows.

There are several minor issues to the manuscript.

1. First, the reference list is pretty slim. The paper is focused on the emergence of Subtype A in Italy from an easter Europe source (i.e., Moldavia/Ukraine). Yest relatively few references are listed relative to the epidemic in Ukraine and none relative to that in Moldavia. They should fix the reference list including all the relevant sources.

We added the requested, as well as other references.

2. Similarly, when they make the discussion on the lower likelihood of HIV sequences from Africa to be grouped in clusters, they should provide references (this is an issue that has been discussed for a long time, starting with Vidal et al JVI 2000)

This point is addressed now in the Discussion section.

3. As the sequences used for the tree construction in Figure 2 are not named, when they list the countries of origin, they should include the reference numbers linked to these sequences to permit the readership to assess the relevance.

We added a supplementary table with requested information

4. I would rename the conclusion section to discussion.

We renamed it.

5. In the discussion related to the possible recombinant nature of the analyzed strains, they should state that the most critical strain (CRF2) has the breaking points in the prot/pol region used for analyses and that there is thus unlikely that the included strains are Ib_NG like.

This point was added in the Discussion section.