Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination

Daiana Mir, Matthieu Jung, Edson Delatorre, Nicole Vidal, Martine Peeters, Gonzalo Bello

Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination

Daiana Mir, Matthieu Jung, Edson Delatorre, Nicole Vidal, Martine Peeters, Gonzalo Bello

PII: S1567-1348(16)30189-7
DOI: doi: 10.1016/j.meegid.2016.05.017
Reference: MEEGID 2750

To appear in:

Received date: 4 February 2016
Revised date: 9 May 2016
Accepted date: 11 May 2016

Please cite this article as: Mir, Daiana, Jung, Matthieu, Delatorre, Edson, Vidal, Nicole, Peeters, Martine, Bello, Gonzalo, Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination, (2016), doi: 10.1016/j.meegid.2016.05.017

TITLE: Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination.

AUTHORS: Daiana Mir[a], Matthieu Jung[b,c,1], Edson Delatorre[a], Nicole Vidal[b], Martine Peeters[b,c] and Gonzalo Bello[a#]

AFFILIATIONS:

[a] Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil.

[b] Unité Mixte Internationale 233, Institut de Recherche pour le Développement, INSERM U1175, and Université Montpellier, Montpellier, France.

[c] Institut de Biologie Computationnelle, LIRMM, UMR 5506 CNRS – Université Montpellier, Montpellier, France.

[1] Current address: IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), INSERM, U596, CNRS, UMR7104, Université de Strasbourg, Illkirch, France.

#Corresponding author:

Gonzalo Bello. Lab. de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ. Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil. Phone: +55 21 3865 8107. Fax: +55 21 3865 8173. E-mail: gbellobr@gmail.com / gbello@ioc.fiocruz.br

ABSTRACT

The HIV-1 CRF02_AG clade is the most prevalent HIV variant in West and West-Central Africa and its detection outside Africa is increasingly common. Little is known, however, about the number and phylodynamics of major CRF02_AG lineages circulating worldwide. To this end, a total of 3,170 HIV-1 CRF02_AG-like *pol* sequences isolated around the world, over a period of 25 years (1989 to 2013), were analyzed using Maximum Likelihood and Bayesian coalescent-based methods. Our results suggest that most of the current CRF02_AG diversity comes from the dissemination of a few founder strains out of Central Africa into West Africa and Cameroon between the late 1960s and the middle 1980s. The CRF02_AG strain introduced into West Africa established a large regional epidemic with low phylogeographic structure. This strain was also successfully disseminated out of the West African region and originated at least three large secondary outbreaks in Cameroon at around the late 1970s, in the former Soviet Union (FSU) countries at around the late 1990s, and in Bulgaria/Germany at around the early 2000s. The CRF02_AG African lineages introduced into Cameroon remained mostly restricted to this country and its neighbors. Demographic reconstructions indicate that major CRF02_AG clades circulating in Africa exhibited a decline in growth rate since the middle 1980s/1990s, whereas CRF02_AG clades in Europe and the FSU countries continue to grow exponentially until the middle to late 2000s. Substantial differences in the median estimated growth rate of the same CRF02_AG clade circulating in different regions (0.63–2.00 $year^{-1}$), and of different CRF02_AG clades circulating in the same country (0.41–0.75 $year^{-1}$) were observed. Thus, the cause of the epidemic outcome of the different HIV-1 CRF02_AG lineages is probably multifactorial.

KEYWORDS: HIV-1; CRF02_AG; Africa; worldwide; phylodynamics.

1. INTRODUCTION

The epidemic dispersion of HIV-1 group M from its location root in Kinshasa, capital of the Democratic Republic of the Congo (DRC), since the first part of the 20th century (Faria et al., 2014) has resulted in the extensive diversity of subtypes, sub-subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) reported across the world. Among the 79 CRFs currently described, CRF02_AG is responsible for the largest number of infections worldwide and is the fourth most prevalent HIV-1 variant accounting for 8% of the global infections (Hemelaar et al., 2011).

The CRF02_AG variant predominates in West and West-Central African countries where it stand for about 50% of the HIV-1 infections (Hemelaar et al., 2011; Lihana et al., 2012); but there is a notable decrease in its frequencies toward Central Africa where it display a prevalence of around 8% of the total infections (Hemelaar et al., 2011; Peeters et al., 2003) and is rarely detected in other African regions. In recent years, there has been an increase in reported sporadic CRF02_AG cases in Europe and North America, mostly caused by migrant flows from endemic regions and global travel (Abecasis et al., 2013; European Centre for Disease Prevention and Control (ECDC), 2014; Hernando et al., 2015; Pyne et al., 2013).

A few indigenous transmission networks of CRF02_AG have been also detected in different regions out of Africa. Autochthonous transmission networks of CRF02_AG have arisen in former Soviet Union (FSU) countries mainly distributed among intravenous drug users (IVDUs), but with increasing prevalence into heterosexual populations (Baryshev et al., 2012; Carr et al., 2005; Eyzaguirre et al., 2007; Kazennova et al., 2014; Laga et al., 2015; Lapovok et al., 2014). Local dissemination of CRF02_AG has been also detected in Brazil, where it has shown the existence of at least two transmission networks with dissemination by both horizontal and vertical

pathways (Delatorre et al., 2015, 2012; Eyer-Silva and Morgado, 2007). Moreover, recent analysis evidence the existence of some native CRF02_AG transmission clusters particularly among HIV-infected men having sex with men (MSM) from France and Belgium and heterosexual population from Switzerland (Brand et al., 2014; Dauwe et al., 2015; Tamalet et al., 2015; Von Wyl et al., 2011).

The field of viral phylodynamics coupled with coalescent-based models has become a powerful tool allowing the recognition of the spatiotemporal dynamics of a variety of viruses and its implementation helped elucidate the origin and dispersion pattern of the CRF02_AG lineage in the Congo River basin (Faria et al., 2012), as well as the demographic dynamics and/or migration routes of CRF02_AG circulating in Guinea Bissau, Cameroon and Brazil (Delatorre et al., 2015; Esbjörnsson et al., 2011; Faria et al., 2012; Véras et al., 2011). The spatiotemporal pattern of dissemination of the CRF02_AG at a global scale, however, remains largely unknown.

The objectives of the present study were to identify and characterize the major HIV-1 CRF02_AG clades circulating in West, West-Central and Central Africa and their dispersion at both regional and global scales. Spatial and temporal information of 3,170 CRF02_AG-like *pol* sequences sampled worldwide over a period of 25 years were used in maximum-likelihood and coalescent-based phylodynamic approaches to determine the prevalence of the major HIV-1 CRF02_AG clades and to reconstruct simultaneously their evolutionary and demographic histories.

2. MATERIALS AND METHODS

2.1. HIV-1 CRF02_AG-like *pol* sequence datasets. A total of 2,246 HIV-1 CRF02_AG-like *pol* sequences, covering the entire protease and partial reverse transcriptase (PR/RT) regions (nucleotides 2,253-3,260 relative to HXB2 genome),

isolated from 20 countries from Central, West-Central and West Africa over a period of 24 years (1990 to 2013) were used in this study (Table S1). Sequences were retrieved from the Los Alamos HIV Database (*n* = 2,113) (http://www.hiv.lanl.gov) and from a local database at the Institut de Recherche pour le Développement, Université Montpellier (*n* = 133). These African sequences were combined with 924 CRF02_AG-like *pol* sequences isolated from 43 countries from the Americas, Europe, Asia and Oceania and with 38 CRF63_02A1-like *pol* sequences isolated in Russia, covering the same genomic region described above and that were available at the Los Alamos HIV Database (Table S1). The subtype assignment of all sequences was confirmed using COMET (Struck et al., 2014) and REGAv3.0 (de Oliveira et al., 2005). Sequences with discordant results were further submitted to Maximum Likelihood (ML) phylogenetic analysis (see bellow) and Bootscan analysis (Lole et al., 1999) with reference samples. All sites with major antiretroviral drug resistance mutations were excluded, leaving 921 nucleotides in the final alignment that is available from authors upon request.

2.2. Phylogenetic analyses. ML trees were inferred with the PhyML v3.0 program (Guindon et al., 2010), under the GTR + I + Γ4 model of nucleotide substitution recommended by the jModeltest program (Posada and Crandall, 2001). The Subtree Pruning and Regrafting (SPR) option was selected as the heuristic tree search method and branch support was estimated with the approximate likelihood-ratio (*a*LRT) SH-like test (Anisimova and Gascuel, 2006). Reference sequences of HIV-1 subtypes B, C, D, F, H, J and K from the Los Alamos HIV Database were used as outgroup. The phylogenetic trees were visualized with FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Major CRF02_AG monophyletic clusters were identified by visual inspection and only those including more than 30 sequences and an *a*LRT score over 0.85 were selected for further analysis.

2.3 Evolutionary and demographic reconstructions. For each CRF02_AG clade identified, the evolutionary rate (μ, units are nucleotide substitutions per site per year, subst./site/year), age of most recent common ancestor ($T_{mrca,}$ years), and mode and rate (r, years^{-1}) of population growth were coestimated by a Bayesian Markov Chain Monte Carlo (MCMC) coalescent-based phylodynamic analyses as implemented in BEAST v1.8 (Drummond and Rambaut, 2007) with BEAGLE (Suchard and Rambaut, 2009) to improve run performance. Since this methodology is computationally prohibitive on large datasets, those identified clades made up by more than 500 sequences were subjected to a sub-sampling strategy (see Supplementary Material for full details of the procedure). Analyses were carried out under the GTR + I + Γ_4 model of nucleotide substitution and a relaxed uncorrelated lognormal molecular clock model (Drummond et al., 2006). A uniform prior was applied on the clock rate (1.5-3.0 x 10^{-3} subs./site/year) on the basis of estimations reported from previous studies (Abecasis et al., 2009). The dynamics of the effective population size (*Ne*) over time were initially estimated by the non-parametric Bayesian skyline plot model (BSP) (Drummond et al., 2005) as coalescent tree prior. Parametric estimates of the growth rates were obtained under three demographic models (exponential, logistic and expansion growth) whose adjustment to the data were assessed using the log marginal likelihood estimation (MLE) based on path sampling (PS) and stepping-stone sampling (SS) approaches (Baele et al., 2012). MCMC were run for 5-50 x 10^7 generations to ensure Effective Sample Size (ESS) values above 200. The ESS and the 95% Highest Probability Density (HPD) values were inspected using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.8 and visualized with FigTree v1.4.2.

3. RESULTS

3.1 Characterization of major HIV-1 CRF02_AG clades circulating in Africa.

The ML phylogenetic analysis performed with 2,246 HIV-1 CRF02_AG-like *pol* sequences from 20 different African countries revealed the existence of five major strongly supported (a*LRT* > 0.85) clades within the radiation of the CRF02_AG (Fig. 1). The clades $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ mostly circulate in Cameroon and together comprise 81% of the CRF02_AG sequences from that country, whereas the clade $CRF02_{WA}$ is the most prevalent one circulating in West Africa and comprises 98% of the CRF02 sequences from that region. Clades $CRF02_{WA}$, $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ were nested among basal sequences from Central Africa and probably represent independent introductions of CRF02_AG strains from Central Africa into West Africa and Cameroon. The clade $CRF02_{CM\text{-}II}$, by contrast, was nested within the $CRF02_{WA}$ clade and probably resulted from a secondary dissemination event of a $CRF02_{WA}$ strain from West Africa into Cameroon.

Analysis of the relative prevalence of major CRF02_AG clades across African countries revealed the existence of four different epidemiologic scenarios (Fig. 2 and Table S2). The CRF02_AG epidemic in Central Africa was mainly composed by basal CRF02_AG lineages (60%), followed by the $CRF02_{WA}$ clade (24%). The CRF02_AG epidemic in West Africa was clearly dominated by the $CRF02_{WA}$ clade (98%). In Gabon and Equatorial Guinea, the $CRF02_{WA}$ clade also predominated (≥ 62%), but a significant fraction of sequences (≥ 14%) branched inside the Cameroonian clades $CRF02_{CM\text{-}I}$ and $CRF02_{CM\text{-}II}$. In Cameroon, the CRF02_AG epidemic was dominated by clades $CRF02_{CM\text{-}I}$ and $CRF02_{CM\text{-}II}$ (71%), and also by a substantial proportion of strains from $CRF02_{WA}$ clade (19%). Despite its variable prevalence (ranging from 19% to 100%), the $CRF02_{WA}$ clade was detected in all African countries analyzed.

3.2. Worldwide dissemination of the major HIV-1 CRF02_AG African clades.

To investigate the role played by each major African CRF02_AG clade in the global dissemination of CRF02_AG, a worldwide set of CRF02_AG-like *pol* gene sequences was subjected to ML phylogenetic analyses, alongside a reference alignment consisting of randomly selected African sequences representative of the major clades. The phylogenetic reconstruction essentially recovered the five major monophyletic CRF02_AG groups (a*LRT* > 0.85) previously identified, with global samples branching within them (Fig. 3). The level of global geographic dispersion observed among the different African CRF02_AG clades varied widely (Fig. 4 and Table S2). The majority (≥ 84%) of the CRF02_AG sequences detected in the Americas, Asia and Europe branched inside the $CRF02_{WA}$ clade. The clades $CRF02_{CM\text{-}I}$ and $CRF02_{CM\text{-}II}$ together comprised between 4% and 17% of the CRF02_AG infections out of Africa, whereas the clades $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ were only detected in Europe and at very low prevalence (< 1%).

The ML phylogenetic reconstruction using the global HIV-1 CRF02_AG-like *pol* sequences also revealed two major non-African strongly supported (a*LRT* ≥ 0.99) monophyletic sub-clades within the $CRF02_{WA}$ radiation (Fig. 3). The sub-clade $CRF02_{BG\text{-}DE}$ included sequences sampled from Bulgaria and Germany between 2006 and 2012, whereas the sub-clade $CRF02_{FSU}$ was composed by sequences from the former Soviet Union (FSU) countries (Russia, Armenia, Kazakhstan, Uzbekistan and Ukraine) isolated between 2002 and 2013. Smaller (n < 20 sequences) country-specific monophyletic sub-clades within the $CRF02_{WA}$ radiation were also detected in others countries (data not shown). These results corroborate the existence of autochthonous transmission networks of CRF02_AG out of Africa that resulted from the introduction and local dissemination of the $CRF02_{WA}$ clade.

3.3. Phylogenetic relationship between $CRF02_{FSU}$ and CRF63_02A1 clades.

The CRF63_02A1 is a HIV-1 variant mainly spreading among IDUs and heterosexual populations from the Russian Federation that was generated by recombination between the CRF02_AG and the subtype A1 clades (Baryshev et al., 2012; Gashnikova et al., 2015; Kazennova et al., 2014; Shcherbakova et al., 2014) This CRF displays a CRF02_AG-like profile in the *pol* gene fragment here selected, which may complicate the subtyping of HIV-1 CRF02_AG-like sequences from FSU countries. To test this, *pol* sequences classified within the $CRF02_{FSU}$ clade were aligned with CRF63_02A1 sequences and with CRF02_AG African sequences representative of the major clades identified in this study. The ML phylogenetic analysis showed that CRF02_AG-like sequences from Armenia, Kazakhstan, Uzbekistan and Ukraine branched at the base of the $CRF02_{FSU}$ clade (Fig. 5). Most Russian sequences, by contrast, are intermixed with CRF63_02A1 viruses in a monophyletic subclade nested within basal $CRF02_{FSU}$ lineages and were thus reclassified as CRF63_02A1-like viruses.

3.4. Timescale and demographic history of CRF02_AG and CRF63_02A1 clades.

The evolutionary and demographic history of major CRF02_AG African clades and of $CRF02_{BG-DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades was reconstructed using a Bayesian coalescent-based approach. The $CRF02_{WA}$ clade was subjected to a sub-sampling strategy because of its large size (n = 1,507), resulting in six subsets (see supplementary material) each of which underwent the same analytical pipeline of the other clades. The median estimated evolutionary rate of the different clades were roughly comparable and all displayed a coefficient of rate variation that did not encompass zero (Tables 1 and S3), thus justifying the use of the relaxed molecular clock model. According to the substitution rates here estimated, the median T_{MRCA} of the different clades dated back to between the late 1960s and the middle 2000s (Tables 1 and S3).

The demographic history estimated through the nonparametric BSP model, suggested that all CRF02_AG clades as well as the CRF63_02A1 clade underwent an initial period of substantial population expansion, followed by a more recent slowdown in its rates of spread (Fig. 6). The growth rate seems to start to decrease between 1985 and 1995 for African CRF02_AG clades, around the middle 2000s for the $CRF02_{FSU}$ clade and around the late 2000s for $CRF02_{BG-DE}$ and CRF63_02A1 clades. To test the significance of such a recent decline in the epidemic growth rate, different parametric growth models were compared for each clade. The logistic growth model provided the best fit to the demographic signal contained in all African CRF02_AG clades and in the CRF63_02A1 clade, whereas the demographic signal contained in the $CRF02_{FSU}$ and $CRF02_{BG-DE}$ clades was nearly equally fitted by both logistic and exponential growth models (Table S4). The median estimated logistic growth rates of major CRF02_AG African clades (0.41 $year^{-1}$ to 0.75 $year^{-1}$) were much lower than those estimated for CRF02_AG and CRF63_02A1 clades circulating in Europe and Asia (1.74 $year^{-1}$ to 2.20 $year^{-1}$) (Tables 1 and S3). According to the exponential growth model, however, the median epidemic growth rates of the $CRF02_{FSU}$ (0.40 $year^{-1}$) and $CRF02_{BG-DE}$ (0.83 $year^{-1}$) clades were lower than those estimated by the logistic one (Table 1).

4. DISCUSSION

The present study embodies a major step toward the identification of the main HIV-1 CRF02_AG lineages circulating worldwide and the characterization of its spatiotemporal dynamics of dissemination. The analyses carried out with 3,170 CRF02_AG-like *pol* sequences sampled around the world support that the current diversity of this HIV-1 variant mostly resulted from the expansion of a few clades with different epidemic outcomes.

The major CRF02_AG African lineage identified here, called $CRF02_{WA}$, probably arose after the introduction of a single founder strain from Central Africa into West Africa. This founder strain was disseminated throughout West Africa since the late 1960s onwards, establishing a large regional epidemic that comprises about 98% of the CRF02_AG sequences from that region here included and displays a very weak geographical structure characterized by country-specific sub-clades of small size ($n < 10$ sequences). The weak phylogeographic structure of the $CRF02_{WA}$ clade reflects multiple and frequent viral exchanges among West African countries that is fully consistent with the strong spatial accessibility (Tatem et al., 2012) and the frequent human mobility (Charrièrre and Fresia, 2008; Gnisci and Trémolieres, 2009) between countries from the West African region, and also coincides with the weak phylogeographic structure observed for other HIV-1 lineages (subtype G and CRF06_cpx) circulating in that region (Delatorre et al., 2014a, 2014b).

Other major CRF02_AG African clades identified here ($CRF02_{CM-I}$, $CRF02_{CM-III}$ and $CRF02_{CM-IV}$), seems to be the result of the expansion of three founder strains probably introduced into Cameroon from Central Africa. The estimated emergence of these CRF02_AG Cameroonian clades encompass a period of almost two decades, ranging from the late 1960s for $CRF02_{CM-I}$ to the middle 1980s for $CRF02_{CM-III}$ and $CRF02_{CM-IV}$. Previous studies described the existence of two (Faria et al., 2012) and three (Véras et al., 2011) major CRF02_AG clades circulating in Cameroon that coincide with clades $CRF02_{CM-I}$/$CRF02_{CM-II}$ and $CRF02_{CM-I}$/$CRF02_{CM-II}$/$CRF02_{CM-III}$ detected in this study, respectively. The greater number of CRF02_AG Cameroonian lineages detected here compared to those previously reported probably arose from the larger number of sequences used in this new study.

The $CRF02_{WA}$ clade was not only successfully disseminated within the West African region, but was also introduced multiple times into other regions of Africa and into other countries around the world, originating a number of secondary outbreaks. The largest secondary $CRF02_{WA}$ outbreaks were detected in Cameroon, Bulgaria/Germany and countries from the FSU, leading to the origin of sub-clades called here $CRF02_{CM-II}$, $CRF02_{BG-DE}$ and $CRF02_{FSU}$, respectively, which were nested within the $CRF02_{WA}$ radiation. The $CRF02_{WA}$ clade and the descendant sub-clades comprise a significant proportion of sequences from West-Central Africa (> 40%), Central Africa (24%), and other regions around the world (> 84%). Thus, the $CRF02_{WA}$ clade is the most successfully disseminated CRF02_AG lineage at a global scale.

The chance of exportation of the CRF02_AG virus from West-Central and Central African regions seems to be much lower than from West Africa. The Cameroonian CRF02_AG clades ($CRF02_{CM-I}$, $CRF02_{CM-II}$, $CRF02_{CM-III}$ and $CRF02_{CM-IV}$) reach a high prevalence in Cameroon (81%) and the neighboring Gabon (38%), but comprise only a minor fraction of the CRF02_AG sequences detected in Angola and DRC (16%), Europe (16%), Equatorial Guinea (14%), America (13%), Asia/Oceania (4%), and West Africa (2%). Similarly, basal CRF02_AG lineages that are prevalent in Angola and DRC (60%) were barely detected outside this region. Given that West Africa hosts a much larger number of CRF02_AG-infected people (~2,500,000) than Cameroon (~330,000) and Central African countries (<100,000) (Hemelaar et al., 2011; Lihana et al., 2012), a more frequent exportation of the $CRF02_{WA}$ clade out of the epicenter like the one supported by our results would be expected.

Demographic reconstructions indicate that major African CRF02_AG clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the middle 1980s-1990s onwards.

The median growth rate of the $CRF02_{WA}$ clade (0.63 $year^{-1}$) was somewhat lower than those previously estimated for subtype G clades and the CRF06_cpx lineage circulating in West Africa (0.75-0.95 $year^{-1}$) (Delatorre et al., 2014a, 2014b) (Fig. S1). Similarly, the median growth rate of different CRF02_AG lineages circulating in Cameroon also varied over a large range (0.41-0.75 $year^{-1}$). Although these results should be interpreted with caution because of the overlap of HPD intervals (Table 1), they support that spatial accessibility may not the only factor that shaped the rate of expansion of the different HIV-1 clades circulating in those African regions. Differences in the onset date of epidemics, transmission dynamics in distinct risk groups and/or viral transmissibility properties might be also responsible for the growth rate variances observed.

Our analyses revealed the existence of two major CRF02_AG transmission networks outside Africa involving individuals from Bulgaria/Germany ($CRF02_{BG-DE}$) and FSU countries ($CRF02_{FSU}$) that probably arose at around the early 2000s and the late 1990s, respectively. Although none of these countries host a large number of West African migrants (Charrière and Fresia, 2008), molecular epidemiologic studies showed that CRF02_AG infections detected in Bulgaria (Ivanov et al., 2013) as well as in Kazakhstan (Eyzaguirre et al., 2007; Lapovok et al., 2014), Uzbekistan (Carr et al., 2005), Kyrgyzstan (Laga et al., 2015), and the Russian Federation (Kazennova et al., 2014) were preferentially associated to IVDUs populations. Given the epidemiological link between the CRF02_AG and IVDUs transmission networks described in the above countries, the origin of clades $CRF02_{BG-DE}$ and $CRF02_{FSU}$ could have been shaped by the rise of international heroin traffic routes linking Afghanistan (the world largest opium producer) to the markets of the Russian Federation and Western Europe (UNODC, 2008).

The HIV-1 epidemic in IVDUs from FSU countries has been mainly driven by a subtype A1 variant characteristic of that region (A_{FSU}), that probably began to spread among IVDUs from Ukraine in the early 1990s and was later disseminated to other FSU countries (Díez-Fuertes et al., 2015). According to our estimations, the $CRF02_{FSU}$ variant began to spread in FSU countries around the late 1990s, probably resulting in a high number of co-infections with the A_{FSU} variant already circulating and the subsequent generation of the $A_{FSU}/CRF02_{FSU}$ recombinant called CRF63_02A1 clade. We estimated the origin of the CRF63_02A1 clade around the middle 2000s, consistent with a previous study (Shcherbakova et al., 2014), supporting a very short time interval (<10 years) between the emergence of the $CRF02_{FSU}$ lineage and the origin of the CRF63_02A1 clade in the IVDUs from FSU countries.

Inspection of the BSP of the $CRF02_{BG-DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades supports a trend toward very recent epidemic stabilization since 2005-2010. The median logistic growth rate estimated for these clades circulating in European and FSU countries were very similar among each other (1.74-2.20 $year^{-1}$) and between three and five times faster than those estimated for African CRF02_AG clades. These extremely fast epidemic growth rates are fully consistent with the preferential dissemination of $CRF02_{BG-DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades through highly connected IVDUs transmission networks, in contrast to the African CRF02_AG clades that are mainly disseminated through heterosexual networks. The exponential demographic model, however, also provide a good fit to the demographic signal in the $CRF02_{FSU}$ and $CRF02_{BG-DE}$ clades and supports lower epidemic expansion rates than those estimated by the logistic model. It is possible that the logistic pattern of the $CRF02_{FSU}$ and $CRF02_{BG-DE}$ clades was more difficult to capture due to its recent stabilization and/or low number of sequences, although we cannot ruled out that those epidemics are still growing exponentially.

5. CONCLUSIONS

In summary, this study reveals that the current CRF02_AG epidemics in West and West-Central African countries resulted from the dissemination of a few founder strains out of Central Africa between the late 1960s and the middle 1980s. The CRF02_AG strain introduced into the West African region ($CRF02_{WA}$) showed a broader geographic dissemination than any other African lineage. Spread of the $CRF02_{WA}$ clade outside Africa lead to the emergence of local transmission networks in Asia and Europe between the late 1990s and the early 2000s. The epidemic outcome of the different CRF02_AG lineages was probably shaped by several factors including: time of origin, spatial accessibility at the epicenter, risk groups transmission dynamics, and viral transmissibility properties.

ACKNOWLEDGMENTS

D.M. was funded by fellowships from “Agencia Nacional de Investigación e Innovación (ANII-Uruguay)” and “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil)”. E.D. was funded by a fellowship from “Programa Nacional de Pós-Doutorado (CAPES-Brazil)”.

REFERENCES

Abecasis, A.B., Vandamme, A.-M., Lemey, P., 2009. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J. Virol. 83, 12917–24. doi:10.1128/JVI.01022-09

Abecasis, A.B., Wensing, A.M., Paraskevis, D., Vercauteren, J., Theys, K., Amc, D., De Vijver, V., Albert, J., Asjö, B., Balotta, C., Beshkov, D., Camacho, R.J., Boucher, C.A., Vandamme, A.-M., 2013. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology 10, 1. doi:10.1186/1742-4690-10-7

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. 55, 539–52. doi:10.1080/10635150600755453

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A. V, 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–67. doi:10.1093/molbev/mss084

Baryshev, P.B., Bogachev, V. V., Gashnikova, N.M., 2012. Genetic characterization of an isolate of HIV type 1 AG recombinant form circulating in Siberia, Russia. Arch. Virol. 157, 2335–2341. doi:10.1007/s00705-012-1442-4

Brand, D., Moreau, A., Cazein, F., Lot, F., Pillonel, J., Brunet, S., Thierry, D., Le Vu, S., Plantier, J.-C., Semaille, C., Barin, F., 2014. Characteristics of patients recently infected with HIV-1 non-B subtypes in France: a nested study within the mandatory notification system for new HIV diagnoses. J. Clin. Microbiol. 52, 4010–6. doi:10.1128/JCM.01141-14

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

Charrière, F., Fresia, M., 2008. West Africa as a Migration and Protection area [WWW Document]. UN High Comm. Refug. URL http://www.refworld.org/docid/4a277db82.html (accessed 1.14.16).

Dauwe, K., Mortier, V., Schauvliege, M., Heuvel, A. Van Den, Fransen, K., Servais, J., Bercoff, D.P., Seguin-devaux, C., Verhofstede, C., 2015. Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate. BMC Infect. Dis. doi:10.1186/s12879-015-1217-0

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E.J., Wensing, A.M.J., van de Vijver, D.A., Boucher,

C.A., Camacho, R., Vandamme, A.-M., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21, 3797–800. doi:10.1093/bioinformatics/bti607

Delatorre, E., Bello, G., Eyer-Silva, W.A.C.-F.S.L., Morgado, M.G., Couto-Fernandez, J.C., 2012. Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil. AIDS Res. Hum. Retroviruses 28, 1369–72. doi:10.1089/aid.2011.0381

Delatorre, E., Mir, D., Bello, G., 2014a. Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa. PLoS One 9, e98908. doi:10.1371/journal.pone.0098908

Delatorre, E., Mir, D., Bello, G., 2014b. Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa. PLoS One 9. doi:10.1371/journal.pone.0098908

Delatorre, E., Velasco-De-Castro, C.A., Pilotto, J.H., Couto-Fernandez, J.C., Bello, G., Morgado, M.G., 2015. Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil. AIDS Res. Hum. Retroviruses 31. doi:10.1089/aid.2015.0183

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

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88. doi:10.1371/journal.pbio.0040088

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–92. doi:10.1093/molbev/msi103

Esbjörnsson, J., Mild, M., Månsson, F., Norrgren, H., Medstrand, P., 2011. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: Origin, demography and migrations. PLoS One 6. doi:10.1371/journal.pone.0017025

European Centre for Disease Prevention and Control (ECDC), 2014. Assessing the burden of key infectious diseases affecting migrant populations in the EU/EEA, Technical Report. doi:10.2900/28792

Eyer-Silva, W. a, Morgado, M.G., 2007. Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. Mem. Inst. Oswaldo Cruz 102, 809–15. doi:S0074-02762007005000112 [pii]

Eyzaguirre, L.M., Erasilova, I.B., Nadai, Y., Saad, M.D., Kovtunenko, N.G., Gomatos, P.J., Zeman, V. V, Botros, B.A., Sanchez, J.L., Birx, D.L., Earhart, K.C., Carr, J.K., 2007. Genetic Characterization of HIV-1 Strains Circulating in Kazakhstan. J. Acquir. Immune Defic. Syndr. 46, 19–23. doi:10.1097/QAI.0b013e318073c620

Faria, N.R., Rambaut, a, Suchard, M. a, Baele, G., Bedford, T., Ward, M.J., Tatem, a J., Sousa, J.D., Arinaminpathy, N., Pepin, J., Posada, D., Peeters, M., Pybus, O.G., Lemey, P., 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science (80-.). 56. doi:10.1126/science.1256739

Faria, N.R., Suchard, M. a., Abecasis, A., Sousa, J.D., Ndembi, N., Bonfim, I., Camacho, R.J., Vandamme, A.M., Lemey, P., 2012. Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. Infect. Genet. Evol. 12, 453–460. doi:10.1016/j.meegid.2011.04.028

Gashnikova, N.M., Bogachev, V. V, Baryshev, P.B., Totmenin, A. V, Gashnikova, M.P., Kazachinskaya, A.G., Ismailova, T.N., Stepanova, S.A., Chernov, A.S., Mikheev, V.N., 2015. A Rapid Expansion of HIV-1 CRF63_02A1 Among Newly Diagnosed HIV-Infected Individuals in the Tomsk Region, Russia. AIDS Res. Hum. Retroviruses 31, 456–460. doi:10.1089/aid.2014.0375

Gnisci, D., Trémolieres, M., 2009. West African Studies, Regional Atlas on West Africa. doi:10.1787/9789264056763-en

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59, 307–321. doi:10.1093/sysbio/syq010

Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25, 679–689. doi:10.1097/QAD.0b013e328342ff93

Hernando, V., Alvarez-Del Arco, D., Alejos, B., Monge, S., Amato-Gauci, A.J., Noori, T., Pharris, A., Del Amo, J., 2015. HIV infection in migrant populations in the European Union and European Economic Area in 2007-2012; an epidemic on the move. J Acquir Immune Defic Syndr 70, 204–211. doi:10.1097/QAI.0000000000000717

Ivanov, I.A., Beshkov, D., Shankar, A., Hanson, D.L., Paraskevis, D., Georgieva, V., Karamacheva, L., Taskov, H., Varleva, T., Elenkov, I., Stoicheva, M., Nikolova, D., Switzer, W.M., 2013. Detailed molecular epidemiologic characterization of HIV-1 infection in Bulgaria reveals broad diversity and evolving phylodynamics. PLoS One 8, e59666. doi:10.1371/journal.pone.0059666

Kazennova, E., Laga, V., Lapovok, I., Glushchenko, N., Neshumaev, D., Vasilyev, A., Bobkova, M., 2014. HIV-1 Genetic Variants in the Russian Far East. AIDS Res.

Hum. Retroviruses 30. doi:10.1089/aid.2013.0194

Laga, V., Lapovok, I., Kazennova, E., Ismailova, A., Beisheeva, N., Asybalieva, N., Glushchenko, N., Bobkova, M., 2015. The Genetic Variability of HIV-1 in Kyrgyzstan□: The Spread of CRF02 _ AG and Subtype A1 Recombinants. J. HIV AIDS 1.2.

Lapovok, I., Kazennova, E., Laga, V., Vasilyev, A., Utegenova, A., Abishev, A., Dzissyuk, N., Tukeev, M., Bobkova, M., 2014. Short communication: molecular epidemiology of HIV type 1 infection in Kazakhstan: CRF02_AG prevalence is increasing in the southeastern provinces. AIDS Res. Hum. Retroviruses 30, 769–774. doi:10.1089/AID.2013.0291

Lihana, R., Ssemwanga, D., Abimiku, A., Ndembi, N., 2012. Update on HIV-1 Diversity in Africa: A Decade in Review. AIDS Rev 14, 83–100.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–60.

Peeters, M., Toure-Kane, C., Nkengasong, J.N., 2003. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. AIDS 17, 2547–2560. doi:10.1097/01.aids.0000096895.73209.89

Posada, D., Crandall, K.A., 2001. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). Mol. Biol. Evol. 18, 897–906.

Pyne, M.T., Hackett, J., Holzmayer, V., Hillyard, D.R., 2013. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. J. Clin. Microbiol. 51, 2662–2669. doi:10.1128/JCM.00880-13

Shcherbakova, N.S., Shalamova, L. a, Delgado, E., Fernández-García, A., Vega, Y., Karpenko, L.I., Ilyichev, A. a, Sokolov, Y. V, Shcherbakov, D.N., Pérez-Álvarez, L., Thomson, M.M., 2014. Short communication: Molecular epidemiology, phylogeny, and phylodynamics of CRF63_02A1, a recently originated HIV-1 circulating recombinant form spreading in Siberia. AIDS Res. Hum. Retroviruses 30, 912–9. doi:10.1089/AID.2014.0075

Struck, D., Lawyer, G., Ternes, A.-M., Schmit, J.-C., Bercoff, D.P., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42, e144. doi:10.1093/nar/gku739

Suchard, M.A., Rambaut, A., 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25, 1370–6. doi:10.1093/bioinformatics/btp244

Tamalet, C., Ravaux, I., Moreau, J., Bré, S., Tourres, C., Richet, H., Abat, C., Colson,

P., 2015. Emergence of Clusters of CRF02_AG and B Human Immunodeficiency Viral Strains Among Men Having Sex With Men Exhibiting HIV Primary Infection in Southeastern France. J. Med. Virol. J. Med. Virol 87. doi:10.1002/jmv.24184

Tatem, A.J., Hemelaar, J., Gray, R.R., 2012. Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa 1–11. doi:10.1097/QAD.0b013e328359a904

UNODC, 2008. Illicit Drug Trends in Central Asia. Available at: www.unodc.org/documents/regional/central-asia/Illicit%20Drug%20Trends_Central%20Asia-final.pdf

Véras, N.M.C., Santoro, M.M., Gray, R.R., Tatem, A.J., Presti, A. Lo, Olearo, F., Cappelli, G., Colizzi, V., Takou, D., Torimiro, J., Russo, G., Callegaro, A., Salpini, R., D'Arrigo, R., Perno, C.-F., Goodenow, M.M., Ciccozzi, M., Salemi, M., 2011. Molecular Epidemiology of HIV Type 1 CRF02_AG in Cameroon and African Patients Living in Italy. AIDS Res. Hum. Retroviruses 27, 1173–1182. doi:10.1089/aid.2010.0333

Von Wyl, V., Kouyos, R.D., Yerly, S., Böni, J., Shah, C., Bürgisser, P., Klimkait, T., Weber, R., Hirschel, B., Cavassini, M., Staehelin, C., Battegay, M., Vernazza, P.L., Bernasconi, E., Ledergerber, B., Bonhoeffer, S., Günthard, H.F., 2011. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. J. Infect. Dis. 204, 1095–1103. doi:10.1093/infdis/jir491

Table 1. Evolutionary and demographic parameters estimated for CRF02_AG and CRF63_02A1 clades.

Clade	N	Sampling interval	Substitution rate (10^{-3})	Coefficient of variation	T_{MRCA}	Growth model	Growth rate
$CRF02_{WA}$*	1,507	1990-2013	1.8 (1.5-2.1)	0.23 (0.19-0.26)	1967 (1961-1974)	LG	0.63 (0.48-0.78)
$CRF02_{CM-I}$	428	1996-2012	1.6 (1.5-1.8)	0.26 (0.22-0.30)	1967 (1962-1973)	LG	0.41 (0.3-0.5)
$CRF02_{CM-II}$	212	1996-2012	1.7 (1.5-2.1)	0.35 (0.29-0.41)	1978 (1972-1984)	LG	0.75 (0.5-1.0)
$CRF02_{CM-III}$	50	2001-2012	1.5 (1.5-1.9)	0.31 (0.16-0.47)	1985 (1979-1989)	LG	0.58 (0.3-1.0)
$CRF02_{CM-IV}$	33	1999-2012	1.6 (1.5-2.0)	0.26 (0.01- 0.44)	1983 (1976-1989)	LG	0.44 (0.2-0.7)
$CRF02_{BG-DE}$	54	2006-2012	1.7 (1.5-2.3)	1.10 (0.70- 1.60)	2003 (1997-2005)	LG	1.74 (0.67-3.59)
						EG	0.84 (0.59-1.11)
$CRF02_{FSU}$	28	2008-2012	1.6 (1.5-1.9)	0.33 (0.07-0.61)	1998 (1996-2000)	LG	2.00 (0.48-4.36)
						EG	0.40 (0.30-0.49)
CRF63_02A1	130	2009-2013	1.6 (1.5-2.0)	0.63 (0.51-0.79)	2004 (2003-2006)	LG	2.10 (1.25-3.25)

* Mean values from all $CRF02_{WA}$ subsets (Table S3). LG: logistic growth. EG: exponential growth.

FIGURE LEGENDS

Figure 1. ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (~1,000 nt) from Central, West-Central and West Africa. Branches are colored according to the geographic origin of each sequence as indicated at the legend (top left). Positions of major CRF02_AG African clades are indicated by shaded boxes. The aLRT support value of each identified clade was > 0.85. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar in the center indicating nucleotide substitutions per site.

Figure 2. Prevalence of $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$, $CRF02_{CM\text{-}IV,}$ $CRF02_{WA}$ clades and basal sequences ($CRF02_{CA}$) among CRF02_AG infected individuals from different African countries, estimated from the phylogenetic analyses presented in Fig. 1. The total number of CRF02_AG sequences analyzed in each country is indicated. The CRF02_AG clades and the countries from each African region are represented by a color code as indicated at the legends at bottom. Two-letter country codes are described in Table S1.

Figure 3. ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (~1,000 nt) isolated around the world alongside reference sequences representative of the major African clades. Branches are colored according to the geographic origin of each sequence as indicated at the legend (top left). Positions of major CRF02_AG clades are indicated by shaded boxes. The aLRT support value of each identified clade was > 0.85. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar in the center indicating nucleotide substitutions per site.

Figure 4. Prevalence of $CRF02_{CM-I}$, $CRF02_{CM-II}$, $CRF02_{CM-III}$, $CRF02_{CM-IV}$ and $CRF02_{WA}$ clades among CRF02_AG infected individuals from different countries out of Africa, estimated from the phylogenetic analyses presented in Fig. 3. The total number of CRF02_AG sequences analyzed in each region is indicated. Each CRF02_AG clade is represented by a color as indicated at the legend. The countries of each region included in the analysis are filled in gray.

Figure 5. ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (~1,000 nt) of the $CRF02_{FSU}$ and the CRF63_02A1 clades alongside reference sequences representative of the major African clades. Branches are colored according to the geographic origin and clade assignment of each sequence as indicated at the legend (top left). Positions of major CRF02_AG and CRF63_02A1 clades are indicated. The aLRT support value of each identified clade was > 0.85. The Cameroonian CRF02_AG clades as well as outgroup sequences were compressed for visual clarity. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar indicating nucleotide substitutions per site.

Figure 6. Demographic history of major HIV-1 CRF02_AG clades and the CRF63_02A1 clade. Median estimates of the effective number of infections using Bayesian skyline (black line) and logistic/exponential growth models (red/blue line) are shown in each graphic together with 95% HPD intervals of the Bayesian skyline estimates (gray area). The vertical axes represent the estimated effective number of infections on a logarithmic scale. Time scale is in calendar years.

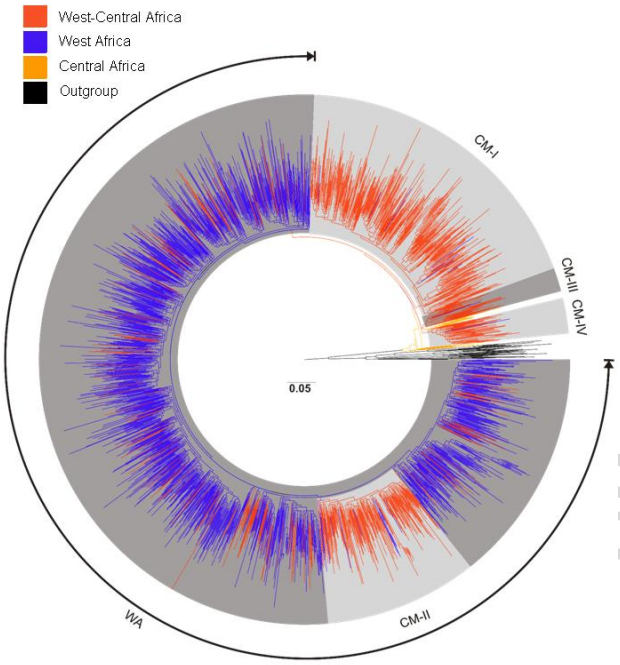

Fig. 1

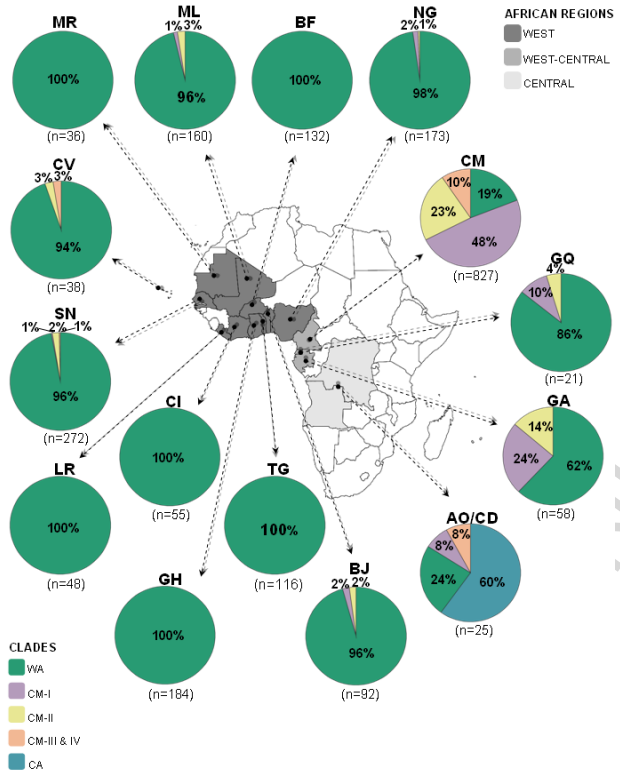

Fig. 2

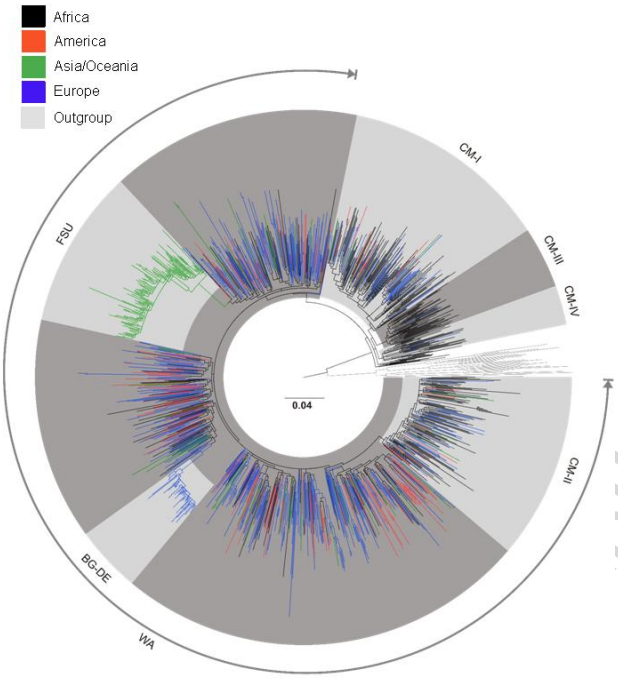

Fig. 3

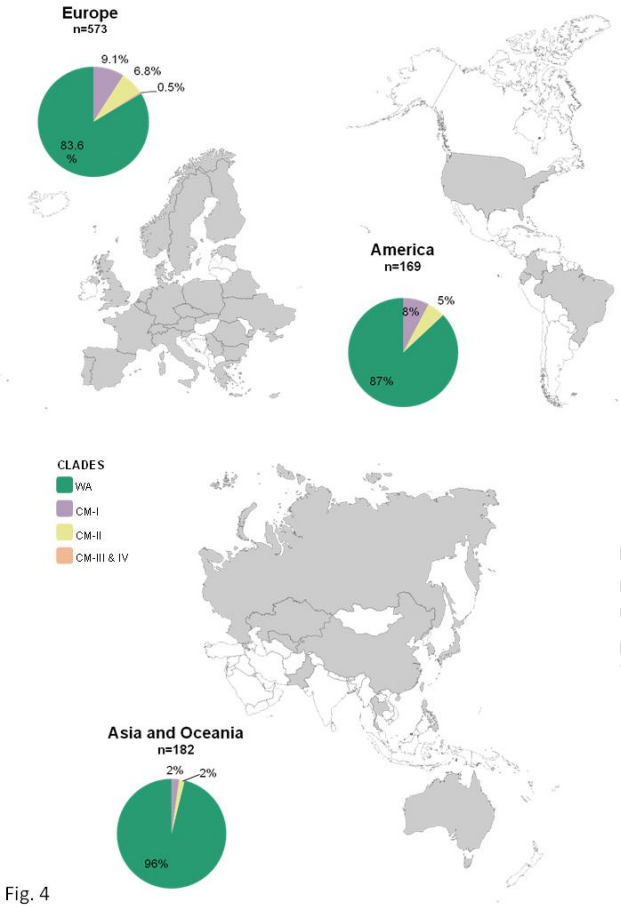

Fig. 4

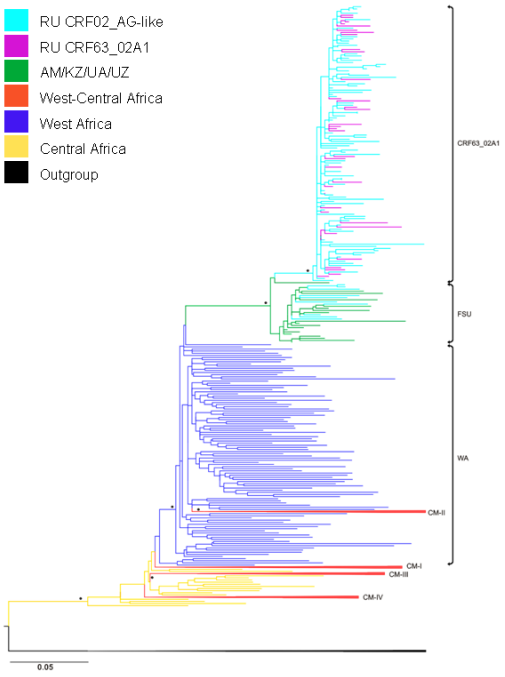

Fig. 5

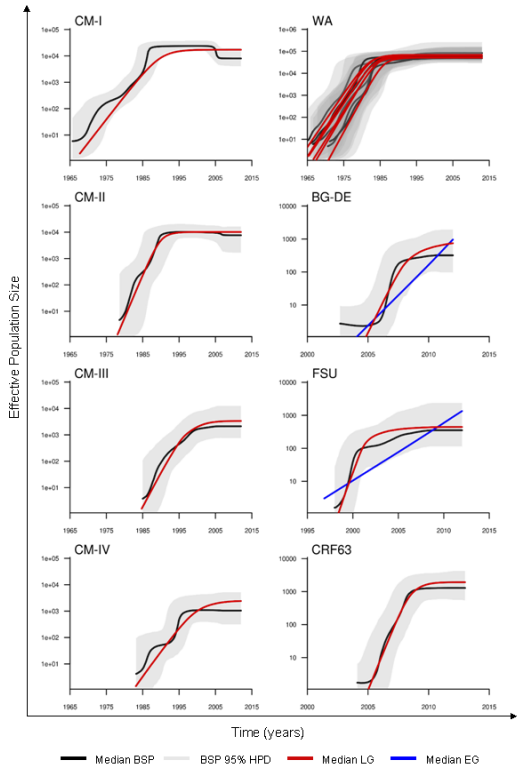

Fig. 6

HIGHLIGHTS

- The current HIV-1 CRF02_AG epidemics in West and West-Central African countries are the result of a few founder strains from Central Africa.
- The West African CRF02_AG clade ($CRF02_{WA}$) showed a high geographic dissemination both inside and outside Africa.
- The introduction of the $CRF02_{WA}$ clade in Asia and Europe lead to the emergence of local epidemics.