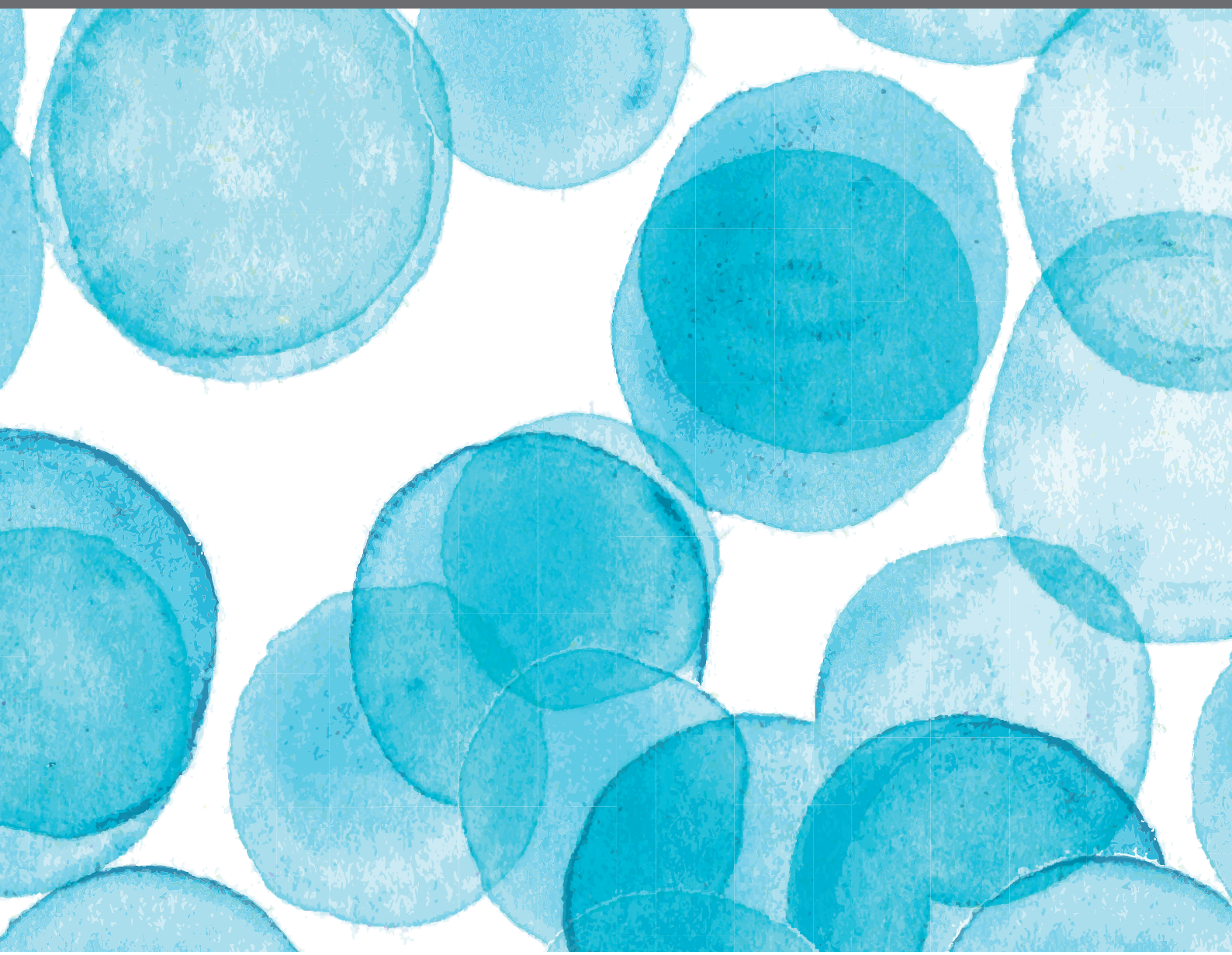
HIV-1 GENETIC DIVERSITY

EDITED BY: Kok Keng Tee, Joris Hemelaar and Michael M. Thomson
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-899-1
DOI 10.3389/978-2-88966-899-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

HIV-1 GENETIC DIVERSITY

Topic Editors:
Kok Keng Tee, University of Malaya, Malaysia
Joris Hemelaar, University of Oxford, United Kingdom
Michael M. Thomson, Instituto de Salud Carlos III (ISCIII), Spain

Citation: Tee, K. K., Hemelaar, J., Thomson, M. M., eds. (2021). HIV-1 Genetic Diversity. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-899-1

Table of Contents

frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 July 2018
doi: 10.3389/fmicb.2018.01738

The HIV-1 Subtype B Epidemic in French Guiana and Suriname Is Driven by Ongoing Transmissions of Pandemic and Non-pandemic Lineages

*Gonzalo Bello[1], Mathieu Nacher[2], Flavia Divino[1], Edith Darcissac[3], Daiana Mir[1] and Vincent Lacoste[3]**

[1] Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [2] Coordination Régionale de la Lutte Contre le VIH (COREVIH) and Centre d'Investigation Clinique – CIC INSERM 1424, Centre Hospitalier de Cayenne "Andrée Rosemon", Cayenne, French Guiana, [3] Laboratoire des Interactions Virus-Hôtes, Institut Pasteur de la Guyane, Cayenne, French Guiana

The HIV-1 subtype B epidemic in French Guiana and Suriname is characterized by the co-circulation of the globally disseminated "$B_{PANDEMIC}$" lineage and of non-pandemic subtype B lineages of Caribbean origin (B_{CAR}). To reconstruct the spatiotemporal pattern of spread of those viral lineages circulating in these two countries, a total of 361 HIV-1 subtype B *pol* sequences recovered from treatment-naive adult patients from French Guiana and Suriname between 2006 and 2012 were combined with $B_{PANDEMIC}$ and B_{CAR} reference sequences. Major Guianese/Surinamese $B_{PANDEMIC}$ and B_{CAR} lineages were identified by Maximum Likelihood phylogenetic analysis and the spatiotemporal and demographic parameters estimated using a Bayesian coalescent-based method. We detected four B_{CAR} and three $B_{PANDEMIC}$ transmission chains of large size that together comprise most pandemic and non-pandemic subtype B sequences from French Guiana ($\geq$52%) and Suriname ($\geq$70%) here analyzed. These major lineages were probably introduced into French Guiana and Suriname from the Caribbean (B_{CAR}) and North/South America ($B_{PANDEMIC}$) between the middle 1970s and the late 1980s and spread among populations from both countries with roughly comparable demographic growth rates. We detected a significant trend for higher viral loads and higher proportion of homosexual/bisexual men among subjects infected with $B_{PANDEMIC}$ relative to B_{CAR} strains in French Guiana. These results show that the HIV subtype B epidemic in French Guiana and Suriname has been driven by multiple active B_{CAR} and $B_{PANDEMIC}$ transmission chains that arose since the middle 1970s onward and operate in both countries simultaneously. Although no significant differences in the epidemic potential of major B_{CAR} and $B_{PANDEMIC}$ lineages were observed, relevant associations between the infecting subtype B lineage and epidemiological and clinical characteristics were detected in French Guiana.

Keywords: HIV-1, subtype B, pandemic, non-pandemic, phylodynamics, French Guiana, Suriname

OPEN ACCESS

Edited by:
Joris Hemelaar, University of Oxford, United Kingdom

Reviewed by:
Cristian Apetrei, University of Pittsburgh, United States
Larance Ronsard, Massachusetts Institute of Technology, United States

****Correspondence:***
Vincent Lacoste
vlacoste@pasteur-cayenne.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *20 April 2018*
Accepted: *11 July 2018*
Published: *31 July 2018*

Citation:
Bello G, Nacher M, Divino F, Darcissac E, Mir D and Lacoste V (2018) The HIV-1 Subtype B Epidemic in French Guiana and Suriname Is Driven by Ongoing Transmissions of Pandemic and Non-pandemic Lineages. Front. Microbiol. 9:1738. doi: 10.3389/fmicb.2018.01738

INTRODUCTION

The Guianas are a region located on the northeastern coast of South America, bordered by Brazil to the south and Venezuela to the west, which includes the French Guiana (an overseas department of France), and the sovereign states of Guyana (known as British Guiana until 1966) and Suriname (part of the Kingdom of the Netherlands until 1975). With a combined population of nearly 1.5 million inhabitants, the Guianas consists of a wide variety of ethnic groups due to historical colonization by Amerindians, Europeans, Africans, and Asians and recent migratory fluxes from neighboring South American and Caribbean countries (Hyles, 2014). In this singular geographic and demographic context, the HIV/AIDS epidemic is a major public health problem and HIV prevalence rates in adult populations from French Guiana, Guyana, and Suriname (1.0–1.5%) are among the highest in the American continent (Nacher et al., 2010; UNAIDS, 2013).

Subtype B is the predominant HIV-1 lineage circulating in French Guiana (Kazanji et al., 2001; Darcissac et al., 2016) and Suriname (Abdoel Wahid et al., 2016); but in sharp contrast to other continental American countries where the epidemic is mostly driven by the globally disseminated "$B_{PANDEMIC}$" lineage, the subtype B epidemic in French Guiana and Suriname is driven by transmission of both $B_{PANDEMIC}$ and of non-pandemic subtype B lineages characteristic of the Caribbean region ("B_{CAR}" lineages) (Cabello et al., 2015). This epidemiological pattern resembles that described in several Caribbean islands (Haiti, the Dominican Republic, Jamaica, The Bahamas and the Lesser Antilles) (Cabello et al., 2014) and in the Northern Brazilian state of Roraima (Divino et al., 2016). Previous phylogenetic analyses revealed that a substantial fraction (30–95%) of subtype B infections in Latin American and Caribbean countries resulted from the expansion of a few local (or regional) $B_{PANDEMIC}$ and B_{CAR} founder strains (Delatorre and Bello, 2013; Cabello et al., 2014, 2015; Mendoza et al., 2014; Mir et al., 2015; Divino et al., 2016), thus supporting a great geographic compartmentalization of the HIV-1 subtype B epidemic in those regions.

Little is known about the spatiotemporal dynamics of dissemination, geographic compartmentalization, and demographic history of the $B_{PANDEMIC}$ and B_{CAR} lineages circulating in French Guiana and Suriname. To answer these questions, we used Maximum Likelihood (ML) and Bayesian coalescent-based methods to analyze a comprehensive data set of 361 HIV-1 subtype B *pol* sequences from French Guiana and Suriname recently described (Abdoel Wahid et al., 2016; Darcissac et al., 2016). The sequences were compared with HIV-1 subtype B *pol* sequences from the Caribbean, South America, North America, and Europe to identify country-specific transmission clusters of the $B_{PANDEMIC}$ and B_{CAR} lineages and to reconstruct their evolutionary and demographic dynamics. We also tested if individuals from French Guiana infected by the $B_{PANDEMIC}$ and B_{CAR} lineages displayed or not comparable epidemiological and clinical characteristics.

MATERIALS AND METHODS

Guianese and Surinamese HIV-1 Subtype B Sequences

HIV-1 subtype B *pol* sequences from treatment-naive adult patients from French Guiana (n = 271) and Suriname (n = 90) recently described (Abdoel Wahid et al., 2016; Darcissac et al., 2016) were included in the present study. HIV-1 sequences were sampled over a time period of seven years (2006–2012) and cover the complete protease (PR) and the first part of the reverse transcriptase (RT) regions (nucleotides 2,253–3,275 of reference strain HXB2). Only one sequence per subject was selected and the subtype of all sequences was confirmed using the REGA HIV subtyping tool v.2 (de Oliveira et al., 2005). HIV-1 *pol* sequences were aligned using the ClustalW program (Thompson et al., 1997) and codons associated with major antiretroviral (ARV) drug resistance positions in PR (n = 12) and RT (n = 21) were excluded. All patients were informed of the possible use of epidemiological and clinical data for research and provided written consent. The project was approved by the Comité de Recherche Clinique (CoRC) Pasteur Institute Paris Project number 2014–2016.

HIV-1 Subtype B Lineage Assignment and Identification of Guianese/Surinamese Subtype B Lineages

HIV-1 subtype B *pol* sequences from French Guiana and Suriname were first aligned with 500 subtype B sequences representative of the $B_{PANDEMIC}$ and the B_{CAR} lineages described previously (Cabello et al., 2014; Mendoza et al., 2014) (**Supplementary Table S1**) and classified within corresponding lineages by using a ML phylogenetic approach. ML trees were inferred with the PhyML program (Guindon et al., 2010) using an online web server (Guindon et al., 2005) under the GTR+I+Γ nucleotide substitution model, as selected by the jModelTest program (Posada, 2008), and the SPR branch-swapping algorithm of heuristic tree search. The reliability of the obtained tree topology was estimated with the approximate likelihood-ratio test (*aLRT*) (Anisimova and Gascuel, 2006) based on the Shimodaira-Hasegawa-like procedure. Trees were rooted using subtype D sequences (the closets HIV-1 group M lineage relative to subtype B) taken from the Los Alamos HIV Database and visualized using the FigTree v1.4.0 program (Rambaut, 2009).

HIV-1 B_{CAR} and $B_{PANDEMIC}$ *pol* sequences from French Guiana and Suriname were next aligned with B_{CAR} sequences from the Caribbean and Brazil and with $B_{PANDEMIC}$ sequences from the United States (US), France, the Netherlands, and Northern Brazil characterized previously (Cabello et al., 2014, 2015, 2016; Divino et al., 2016) (**Supplementary Table S2**). Sequences were subjected to ML analyses as described above and Guianese and Surinamese B_{CAR} and $B_{PANDEMIC}$ transmission clusters were defined as highly supported ($aLRT \geq 0.85$) monophyletic clusters mostly (>90%) composed

by sequences from these countries. For putative intra-subtype B_{CAR}/$B_{PANDEMIC}$ recombinant sequences, similarity plots depicting the percentage identity to a panel of $B_{PANDEMIC}$, B_{CAR}, and subtype D reference sequences were generated using SimPlot v.3.5.1 (Lole et al., 1999) and Neighbor-Joining phylogenetic trees of different *pol* gene fragments were reconstructed under the Tamura-Nei model, in 500 bootstrapped datasets, using MEGA v6 (Tamura et al., 2013).

Spatiotemporal and Demographic Reconstructions

To reconstruct spatiotemporal dynamics and identify the most probable source location of major B_{CAR} Guianese/Surinamese lineages here identified, we selected B_{CAR} sequences from the major Caribbean islands with high prevalence of non-pandemic strains and B_{CAR} sequences from neighboring South American countries, including: all B_{CAR} sequences from Hispaniola and sequences of major B_{CAR} lineages circulating in Trinidad and Tobago ($B_{CAR\text{-}TT}$), Jamaica ($B_{CAR\text{-}JM\text{-}I}$), Brazil ($B_{CAR\text{-}BR\text{-}I}$ and $B_{CAR\text{-}BR\text{-}II}$), and Guyana ($B_{CAR\text{-}GY}$) identified in previous studies (Cabello et al., 2014, 2015; Divino et al., 2016) (**Supplementary Table S3**). Similarly, to identify the most probable source location of major $B_{PANDEMIC}$ Guianese/Surinamese lineages we selected a subset of $B_{PANDEMIC}$ reference sequences from regions with the highest human flux from/to French Guiana and Suriname (the Caribbean, South America, Central America, North America, and Europe). A total of 40 $B_{PANDEMIC}$ reference sequences from each geographic region with known date of isolation and with the highest similarity to Guianese and Surinamese sequences were selected using the basic local alignment search tool (BLAST)[1] (**Supplementary Table S4**).

The evolutionary rate, the age of the most recent common ancestor (T_{MRCA}, years), the spatial diffusion pattern and the rate (r, years^{-1}) of population growth of major HIV-1 Guianese/Surinamese subtype B lineages were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 (Drummond et al., 2002; Drummond and Rambaut, 2007) with BEAGLE (Suchard and Rambaut, 2009) to improve run-time. Because regression analyses using program TempEst (Rambaut et al., 2016) revealed that subtype B *pol* datasets here compiled does not contain sufficient temporal signal for reliable time-scale estimations (X-intercept [TMRCA] < 1,910), Bayesian MCMC analyses were performed using the GTR+I+$\tilde{A}_4$ nucleotide substitution model and a relaxed uncorrelated lognormal molecular clock model (Drummond et al., 2006) with a uniform prior distribution on the substitution rate that encompass mean values previously estimated for the subtype B *pol* gene ($2.0–3.0 \times 10^{-3}$ subst./site/year) (Hue et al., 2005; Zehender et al., 2010; Chen et al., 2011; Mendoza et al., 2014). Migration events throughout the phylogenetic histories were reconstructed using a reversible discrete phylogeography model (Lemey et al., 2009) with a CTMC rate reference prior (Ferreira and Suchard, 2008). Changes in effective population size through time for each major HIV-1 Guianese/Surinamese subtype B lineages was independently estimated using a Bayesian Skyline coalescent tree prior (Drummond et al., 2005). Estimates of the population growth rate were obtained using the parametric model (logistic, exponential, or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood estimation based on path sampling (PS) and stepping-stone sampling (SS) methods (Baele et al., 2012). MCMC chains were run for $50–200 \times 10^6$ generations. Convergence (Effective Sample Size > 200) and uncertainty (95% Highest Probability Density [HPD] values) in parameter estimates were assessed using the TRACER v1.6 program (Rambaut and Drummond, 2007). Maximum clade credibility (MCC) trees were summarized with TreeAnnotator v1.7.5 and visualized with FigTree v1.4.0.

Estimation of HIV Incidence Temporal Trend in French Guiana

To estimate the HIV incidence in French Guiana we used the Spectrum v5.51 package[2]. AIM (AIDS Impact Model) and CSAVER (Case Surveillance and Vital Registration) incidence fitting tools were used with a start in 1970 and projections until 2013. Historical HIV programmed data, treatment eligibility criteria for adults and children for different periods, the proportion of pregnant women with access to prevention of mother-to-child transmission of HIV, the number of patients receiving ARV therapy, the median CD4-count at ARV initiation, the proportion of virologically suppressed treated patients and the proportion of lost to follow-up patients each year were entered in Spectrum. The Epidemic was modeled as a concentrated epidemic. The yearly progression rate to the next CD4 category and the HIV mortality with and without ARV were selected from the options in Spectrum based on Latin America and the Caribbean. The default ratio of fertility of infected women versus uninfected ones was used.

Statistical Analyses

Epidemiological and demographic characteristics of the cohort included in the present study were compared using Fisher's exact test or chi2 implemented in Stata 13 software. Statistical significance was defined as $p < 0.05$.

RESULTS

Identification of Major Guianese/Surinamese Subtype B Lineages

HIV-1 subtype B sequences from French Guiana and Suriname (n = 361) were combined with viral strains representative of the $B_{PANDEMIC}$ (n = 300) and B_{CAR} (n = 200) lineages (**Supplementary Table S1**). The ML analysis revealed that, as expected, a very large proportion of HIV-1 subtype B sequences from French Guiana (60%) and Suriname (50%) were intermixed among B_{CAR} strains; whereas the others branched

[1] www.ncbi.nlm.nih.gov/BLAST

[2] http://www.avenirhealth.org/software-spectrum.php

within the well-supported (*aLRT* = 0.90) $B_{PANDEMIC}$ lineage (**Supplementary Figure S1** and **Supplementary Table S5**). A few sequences from French Guiana (4%) and Suriname (6%) remained unclassified as they branched basal to the $B_{PANDEMIC}$ lineage, but were not intermixed among B_{CAR} sequences and, when included, greatly reduced the support of the $B_{PANDEMIC}$ lineage (*aLRT* < 0.80) (data not shown). Similarity plots revealed that most unclassified sequences displayed a higher similarity to B_{CAR} reference strains in the initial portion of the *pol* fragment (covering the entire PR) compared to $B_{PANDEMIC}$ strains, and a higher similarity to $B_{PANDEMIC}$ sequences in the final portion of the *pol* fragment (**Supplementary Figure S2**). This suggests that they may represent intra-subtype $B_{CAR}/B_{PANDEMIC}$ recombinant sequences. The phylogenetic branching pattern at the initial and final portions of the *pol* fragment also supports a $B_{CAR}/B_{PANDEMIC}$ recombinant structure for half of unclassified sequences (**Supplementary Figure S2**). However, this result should be interpreted with caution due the overall low support (bootstrap < 60%) at nodes of phylogenetic trees.

The B_{CAR} and $B_{PANDEMIC}$ sequences from French Guiana and Suriname were then combined with B_{CAR} and $B_{PANDEMIC}$ sequences of countries from the Americas (Caribbean region and Brazil) and Europe (France and the Netherlands) that maintain historical intense migratory fluxes with French Guiana and Suriname. The ML phylogenetic analyses revealed that most B_{CAR} and $B_{PANDEMIC}$ sequences from French Guiana (≥52%) and Suriname (≥70%) branched within a few monophyletic lineages of large size (n > 10 sequences) that were exclusively composed of sequences from both countries ($B_{CAR\text{-}GF/SR\text{-}I}$, $B_{CAR\text{-}GF/SR\text{-}II}$, $B_{CAR\text{-}GF/SR\text{-}III}$, $B_{PAN\text{-}GF/SR\text{-}I}$, and $B_{PAN\text{-}GF/SR\text{-}II}$), from French Guiana ($B_{PAN\text{-}GF\text{-}I}$), or that also include sequences from neighboring countries ($B_{CAR\text{-}SA\text{-}I}$) (**Figure 1** and **Supplementary Table S6**). Geographic distribution of major Guianese/Surinamese B_{CAR} and $B_{PANDEMIC}$ lineages was not homogenous across countries (**Supplementary Tables S7**, **S8**). The remaining B_{CAR} and $B_{PANDEMIC}$ sequences from French Guiana and Suriname branched in country-specific lineages of small size (≤10%) or appeared as non-clustered infections (≤41%) (**Supplementary Table S6**).

Spatiotemporal Dissemination of Major Guianese/Surinamese Subtype B Lineages

To identify the most probable source location of major B_{CAR} and $B_{PANDEMIC}$ Guianese/Surinamese subtype B lineages, those sequences were combined with sequences of major B_{CAR} lineages circulating in the Caribbean and Brazil and with $B_{PANDEMIC}$ sequences from the Americas and Europe with the highest BLAST search similarity score to Guianese/Surinamese sequences. Bayesian phylogeographic analyses support that lineage $B_{CAR\text{-}SA\text{-}I}$ arose after dissemination of a single variant strain from Trinidad and Tobago (*Posterior State Probability* [*PSP*] = 0.98) into French Guiana (*PSP* = 0.93) at around the middle 1970s, that subsequently passed the $B_{CAR\text{-}SA\text{-}I}$ lineage to Suriname, Brazil, and Guyana at multiples times, originating the sublineages $B_{CAR\text{-}BR\text{-}I}$ and $B_{CAR\text{-}GY}$ (**Figure 2A** and **Supplementary Figure S3**). The other Guianese/Surinamese B_{CAR} lineages most probably arose after dissemination of viral strains from Hispaniola (*PSP* ≥ 0.97) into Suriname (lineage $B_{CAR\text{-}GF/SR\text{-}I}$, *PSP* = 0.68) and French Guiana (lineages $B_{CAR\text{-}GF/SR\text{-}II}$ and $B_{CAR\text{-}GF/SR\text{-}III}$, *PSP* ≥ 0.97) between the late 1970s and the middle 1980s, followed by multiple viral exchanges between those countries (**Figure 2A** and **Table 1**). The lineages $B_{PAN\text{-}GF/SR\text{-}I}$ and $B_{PAN-GF/SR\text{-}II}$ most probably arose after concurrent dissemination of $B_{PANDEMIC}$ strains from North America (*PSP* = 1) into French Guiana (*PSP* = 0.97) and Suriname (*PSP* = 0.75), respectively, at around the middle 1980s and were later disseminated between French Guiana and Suriname at multiple times (**Figure 2B** and **Table 1**). We detected only sporadic disseminations of the $B_{PAN\text{-}GF/SR\text{-}I}$ lineage from French Guiana into Northern Brazil (Amapá), North America (the US), and Europe (the Netherlands). The $B_{PAN\text{-}GF-I}$ lineage most probably arose after spreading of a $B_{PANDEMIC}$ strain from South America (*PSP* = 0.92) into French Guiana (*PSP* = 0.99) at around the early 1990s, with no evidence of dissemination outside this country (**Figure 2B** and **Table 1**).

Demographic History of Major Guianese/Surinamese Subtype B Lineages

Bayesian coalescent inference was next used to reconstruct the population dynamics of major HIV-1 subtype B Guianese/Surinamese lineages with more than 30 sequences. The Bayesian skyline plot (BSP) analyses suggest that lineages $B_{CAR\text{-}SA\text{-}I}$, $B_{CAR\text{-}GF/SR-I}$, and $B_{PAN\text{-}GF/SR\text{-}I}$ experienced an initial phase of slow growth until 1985–1990, followed by a phase of exponential growth during the 1990s and subsequent decline in growth rate from the middle 1990s and the middle 2000s (**Figures 3A–C**). The estimated temporal change of the HIV incidence rate in French Guiana also supports an epidemic stabilization around the early 2000s; but points to a continuous reduction of the HIV incidence afterward that was not captured by our coalescent-based demographic inference (**Figure 3D**). The logistic demographic model fit the data better than the other models tested (log BF > 3) for lineages $B_{CAR\text{-}SA\text{-}I}$ and $B_{CAR\text{-}GF/SR\text{-}I}$; but was modestly supported (log BF = 1.9) over the exponential one for lineage $B_{PAN\text{-}GF/SR\text{-}I}$, probably due to the very recent stabilization phase (**Supplementary Table S9**). According to the logistic growth model, the mean estimated epidemic growth rates of lineages $B_{CAR\text{-}SA\text{-}I}$ (0.46 $year^{-1}$), $B_{PAN\text{-}GF/SR\text{-}I}$ (0.45 $year^{-1}$), and $B_{CAR\text{-}GF/SR\text{-}I}$ (0.30 $year^{-1}$) were roughly comparable and corresponded to mean R_0 of between 3.4 and 4.7 (**Table 1**).

Association Between Epidemiological Characteristics and HIV-1 Subtype B Lineages

Analysis of the epidemiological characteristics of Guianese subjects infected with $B_{PANDEMIC}$ and B_{CAR} lineages revealed that both viral lineages circulated among males and females of different age groups. Most subjects infected with both

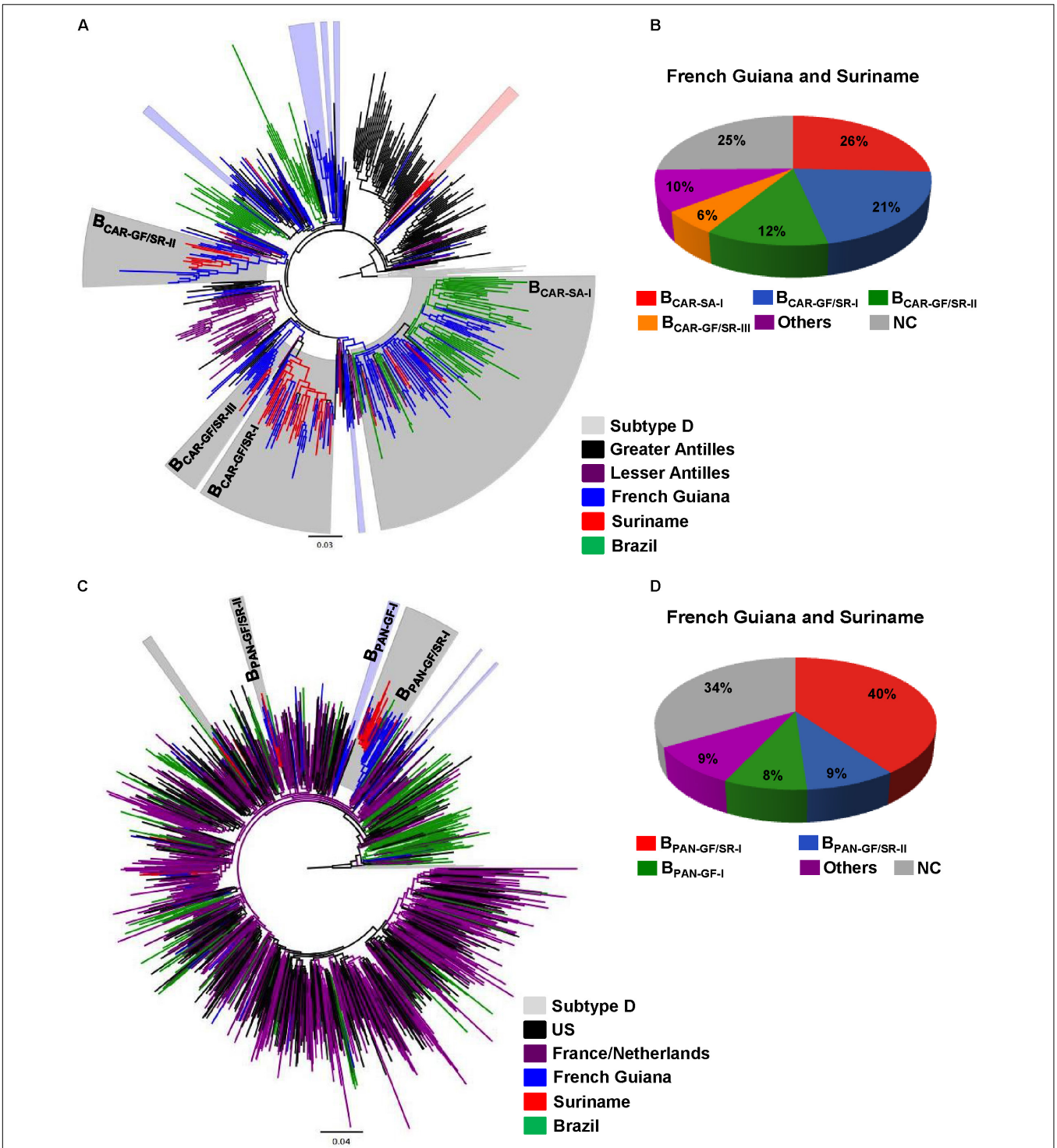


FIGURE 1 | Identification of major Guianese/Surinamese subtype B lineages. **(A,C)** ML phylogenetic trees of HIV-1 B_{CAR} and $B_{PANDEMIC}$ *pol* sequences circulating in French Guiana and Suriname together with representative B_{CAR} sequences from the Caribbean and Brazil and $B_{PANDEMIC}$ sequences from the US, France, the Netherlands and Northern Brazil. Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). Shaded boxes highlight the position of B_{CAR} and $B_{PANDEMIC}$ highly supported (SH-*aLRT* $\geq$ 0.85) clusters mostly/exclusively composed by Guianese and/or Surinamese sequences. Major ($n \geq 10$) B_{CAR} and $B_{PANDEMIC}$ lineages detected in French Guiana and Suriname are indicated with names. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. **(B,D)** Estimated proportion of HIV-1 sequences branching in major and minor Guianese/Surinamese clusters as well of non-clustered (NC) sequences among B_{CAR} and $B_{PANDEMIC}$ infected individuals from French Guiana and Suriname according to the ML analyses.

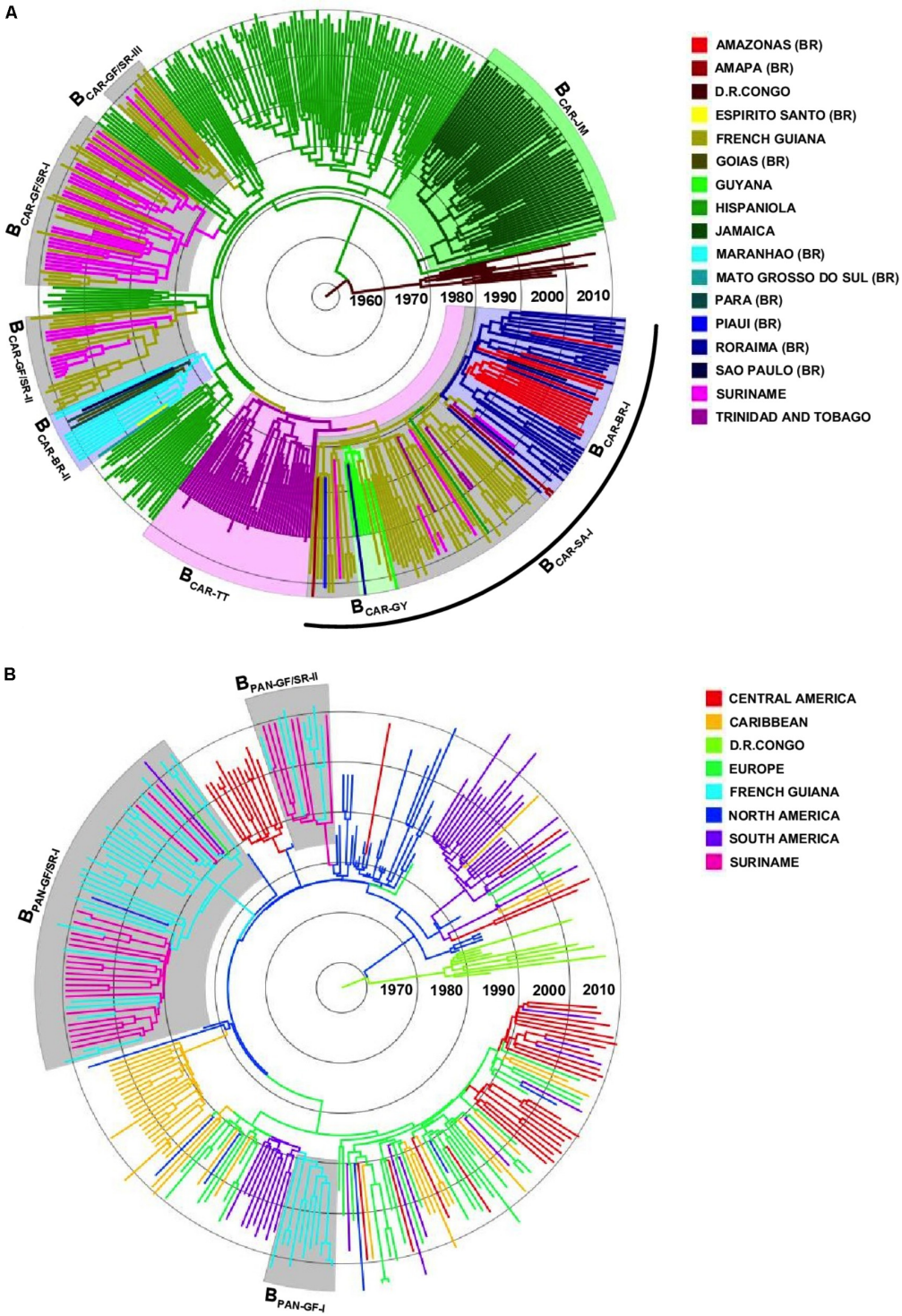


FIGURE 2 | Spatiotemporal dissemination of major Guianese/Surinamese subtype B lineages. **(A)** Time-scaled Bayesian phylogenetic tree of HIV-1 B_{CAR} *pol* sequences from French Guiana, Suriname, Brazil, Guyana, and the Caribbean combined with subtype D reference sequences. Shaded boxes highlight the position of major B_{CAR} lineages from French Guiana/Suriname, Trinidad and Tobago, Jamaica, Brazil, and Guyana. The arc indicates the position of the $B_{CAR\text{-}SA\text{-}I}$ lineage. **(B)** Time-scaled Bayesian phylogenetic tree of HIV-1 $B_{PANDEMIC}$ *pol* sequences from the French Guiana, Suriname and closely related $B_{PANDEMIC}$ sequences from the Caribbean, South/Central/North America, and Europe, combined with subtype D reference sequences. Shaded boxes highlight the position of major $B_{PANDEMIC}$ lineages circulating in French Guiana and Suriname. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend at right. Branch lengths are drawn to a scale of years. The trees are automatically rooted under the assumption of a relaxed molecular clock.

TABLE 1 | Phylogeographic, evolutionary, and demographic parameters estimated for major HIV-1 B_{CAR} and $B_{PANDEMIC}$ lineages circulating in French Guiana and Suriname.

Clade	*N*	Sampling interval	Origin (*PSP*)	T_{MRCA} (95% HPD)	Growth model	Growth rate (95% HPD)	R_0 (95% HPD)
$B_{CAR-SA-I}$	54	2000–2012	GF (0.93)	1977 (1973–1981)	Logistic	0.46 (0.30–0.64)	4.7 (3.4–6.1)
$B_{CAR-GF/SR-I}$	45	2000–2012	SR (0.68)	1978 (1974–1982)	Logistic	0.30 (0.21–0.40)	3.4 (2.7–4.2)
$B_{CAR-GF/SR-II}$	25	2007–2012	GF (0.97)	1980 (1975–1985)	–	–	–
$B_{CAR-GF/SR-III}$	12	2007–2011	GF (1.00)	1984 (1979–1988)	–	–	–
$B_{PAN-GF/SR-I}$	55	2007–2012	GF (0.97)	1985 (1982–1988)	Logistic	0.45 (0.27–0.70)	4.6 (3.2–6.6)
$B_{PAN-GF/SR-II}$	13	2007–2011	SR (0.75)	1987 (1983–1990)	–	–	–
$B_{PAN-GF-I}$	11	2006–2011	GF (0.99)	1990 (1987–1992)	–	–	–

GF, French Guiana; SR, Suriname.

$B_{PANDEMIC}$ and B_{CAR} lineages were mainly followed-up at clinics from the capital city Cayenne and born outside French Guiana (**Table 2**). Although heterosexual sex was the main mode of HIV transmission for both lineages, the proportion of homosexual/bisexual men among those infected by the $B_{PANDEMIC}$ lineage (10%) was five-times higher than among those infected by B_{CAR} lineages (2%). Among clinical characteristics, significant differences were observed in subject distribution according to plasma RNA viral load. The proportion of subjects with high viral loads (>10,000 copies/ml) among those infected with the $B_{PANDEMIC}$ lineage (64%) was higher that among those infected with B_{CAR} lineages (46%).

Significant differences in epidemiological and clinical variables were also observed among major Guianese/Surinamese subtype B transmission chains (**Supplementary Table S10**). In contrast to the overall pattern, most (65%) subjects infected with the lineage $B_{CAR-GF/SR-I}$ were followed-up at clinics from Saint Laurent du Maroni, the second most populous city of French Guiana located at the border with Suriname. Significant associations were observed between infecting strain and the country of origin, including: born in Guyana and lineage $B_{CAR-SA-I}$, born in Suriname and lineage $B_{CAR-GF/SR-I}$, and born in mainland France and lineage $B_{CAR-GF/SR-II}$. Notably, individuals of Haiti origin represent a large proportion (20%) of subjects from French Guiana infected with B_{CAR} strains, but a low fraction (5%) of those infected with major Guianese/Surinamese B_{CAR} lineages. We also observed that the proportion of subjects with high viral loads among those infected with lineage $B_{PAN-GF/SR-I}$ (82%) was higher than among those infected with lineages $B_{CAR-SA-I}$ (43%) and $B_{CAR-GF/SR-II}$ (25%).

DISCUSSION

This study revealed that the HIV-1 subtype B epidemic in French Guiana and Suriname has been driven by multiple active B_{CAR} and $B_{PANDEMIC}$ transmission chains that arose from several independent sources and operate in both countries simultaneously. Most B_{CAR} (≥59%) and $B_{PANDEMIC}$ (≥40%) sequences from French Guiana and Suriname here analyzed branched within regional-specific monophyletic lineages that comprise sequences from both countries; while only a minor fraction of B_{CAR} (≤10%) and $B_{PANDEMIC}$ (≤18%) sequences branched within country-specific clusters containing only sequences from a single country. This high degree of phylogenetic intermixing of HIV sequences is fully consistent with the intense cross-border population mobility between French Guiana and Suriname (Collomb and Fouck, 2016; Jaries et al., 2017) and supports the need to develop a coordinated bi-national healthcare response.

Our analyses identify four major B_{CAR} (designated as $B_{CAR-SA-I}$ and $B_{CAR-GF/SR-I}$ to $B_{CAR-GF/SR-III}$) and three major $B_{PANDEMIC}$ (designated as $B_{PAN-GF/SR-I}$, $B_{PAN-GF/SR-II}$, and $B_{PAN-GF-I}$) transmission chains that together accounted for 54 and 73% of all HIV-1 subtype B sequences from French Guiana and Suriname here analyzed, respectively. This resembles the estimated proportion of HIV-1 subtype B infections that resulted from the expansion of major local (country- and regional-specific) $B_{PANDEMIC}$ and/or B_{CAR} lineages in Argentina (31%), Brazil (31%), Mexico (37%), El Salvador (41%), Peru (51%), Jamaica (53%), Cuba (70%), Panama (77%), Honduras (91%), and Trinidad and Tobago (94%) (Delatorre and Bello, 2013; Cabello et al., 2014; Mendoza et al., 2014; Mir et al., 2015). This level of geographic compartmentalization is much higher than that observed for the subtype B epidemic in Europe where countries are highly interconnected and transcontinental migration is a significant driving force of virus dispersal (Paraskevis et al., 2009; Magiorkinis et al., 2016) and supports that subtype B transmissions in Latin America and the Caribbean are mainly occurring between individuals from the same country or neighboring countries (Junqueira et al., 2016).

The lineage $B_{CAR-SA-I}$ was the most widely disseminated B_{CAR} lineage in French Guiana and the northern South America region. According to our Bayesian phylogeographic analysis, this regional lineage arose after dissemination of a B_{CAR-TT} strain from Trinidad and Tobago into French Guiana at around the middle 1970s and later spread to Suriname, Guyana, and Northern Brazilian region. Nevertheless, a significant proportion (10%) of immigrants residing in Trinidad and Tobago is from Guyana (Borland et al., 2004). In addition, a very high proportion (45%) of subjects infected with the lineage $B_{CAR-SA-I}$ in French Guiana where from Guyana. At last, the intense human mobility across the Roraima border with Guyana (Pereira, 2006; Corbin, 2007) coincides with a high

TABLE 2 | Epidemiological information of subjects from French Guiana infected by HIV-1 B_{CAR} and $B_{PANDEMIC}$ clades.

Characteristic	Total (*n* = 260)	B_{CAR} (*n* = 162)	$B_{PANDEMIC}$ (*n* = 98)	*P*
Sampling interval (years)	2007–2012	2007–2012	2007–2012	–
HIV diagnosis*				
1980–1999	17 (7%)	10 (6%)	7 (7%)	0.67
2000–2005	35 (13%)	25 (15%)	10 (10%)	
2006–2012	202 (78%)	123 (76%)	79 (81%)	
Unknown	6 (3%)	4 (2%)	2 (2%)	
Age group (years)**				
18–24	23 (9%)	14 (9%)	9 (9%)	0.55
25–34	89 (34%)	56 (35%)	33 (34%)	
35–44	72 (28%)	49 (30%)	23 (23%)	
>44	76 (29%)	43 (26%)	33 (34%)	
Sex**				
Male	113 (43%)	65 (40%)	48 (49%)	0.16
Female	147 (57%)	97 (60%)	50 (51%)	
Mode of transmission*				
Homosexual/Bisexual	14 (5%)	4 (2%)	10 (10%)	**0.009**
Heterosexual	229 (88%)	150 (93%)	79 (81%)	
Others	1 (<1%)	0	1 (1%)	
Unknown	16 (6%)	8 (5%)	8 (8%)	
Geographic location*				
Cayenne	193 (74%)	123 (76%)	70 (71%)	0.76
Saint Laurent du Maroni	54 (21%)	31 (19%)	23 (23%)	
Others	4 (2%)	3 (2%)	1 (1%)	
Unknown	9 (3%)	5 (3%)	4 (4%)	
Country of birth**				
French Guiana	31 (12%)	18 (11%)	13 (13%)	0.25
Haiti	56 (22%)	32 (20%)	24 (24%)	
Suriname	52 (20%)	32 (20%)	20 (20%)	
Guyana	33 (13%)	28 (17%)	5 (5%)	
France	33 (13%)	21 (13%)	12 (12%)	
Brazil	27 (10%)	15 (9%)	12 (12%)	
Others	13 (5%)	7 (4%)	6 (6%)	
Unknown	15 (6%)	9 (6%)	6 (6%)	
Clinical condition*				
Asymptomatic (A)	194 (75%)	119 (73%)	75 (76%)	0.57
Symptomatic (B)	24 (9%)	18 (11%)	6 (6%)	
AIDS (C)	33 (13%)	20 (12%)	13 (13%)	
Unknown	9 (3%)	5 (3%)	4 (4%)	
Viral load (copies/ml)*				
<LD	10 (4%)	7 (4%)	3 (3%)	**0.03**
51–2,000	42 (16%)	32 (20%)	10 (10%)	
2,001–10,000	70 (27%)	48 (30%)	22 (22%)	
>10,000	138 (53%)	75 (46%)	63 (64%)	
CD4 count (cells/ml)**				
350–500	133 (51%)	85 (52%)	48 (49%)	0.58
>500	127 (49%)	77 (48%)	50 (51%)	

**Fisher's exact test. ***Pearson's chi2. P values < 0.05 are in bold.*

prevalence of the lineage $B_{CAR\text{-}SA\text{-}I}$ in that Northern Brazilian state (Divino et al., 2016). These epidemiological data therefore suggest that lineage $B_{CAR\text{-}SA\text{-}I}$ arose in Guyana. Nevertheless, the scarcity of HIV-1 sequences from that country seriously constrained our phylogeographic inference. The other major Guianese/Surinamese B_{CAR} lineages most probably arose in the island of Hispaniola and were introduced into French Guiana ($B_{CAR\text{-}GF/SR\text{-}II}$ and $B_{CAR\text{-}GF/SR\text{-}III}$) or Suriname ($B_{CAR\text{-}GF/SR\text{-}I}$) between the late 1970s and the middle 1980s. It is interesting to note that a significant proportion (20%) of B_{CAR}-infected subjects living in French Guiana were born in Haiti but individuals of Haitian origin only represent a minor fraction (5%) of lineages

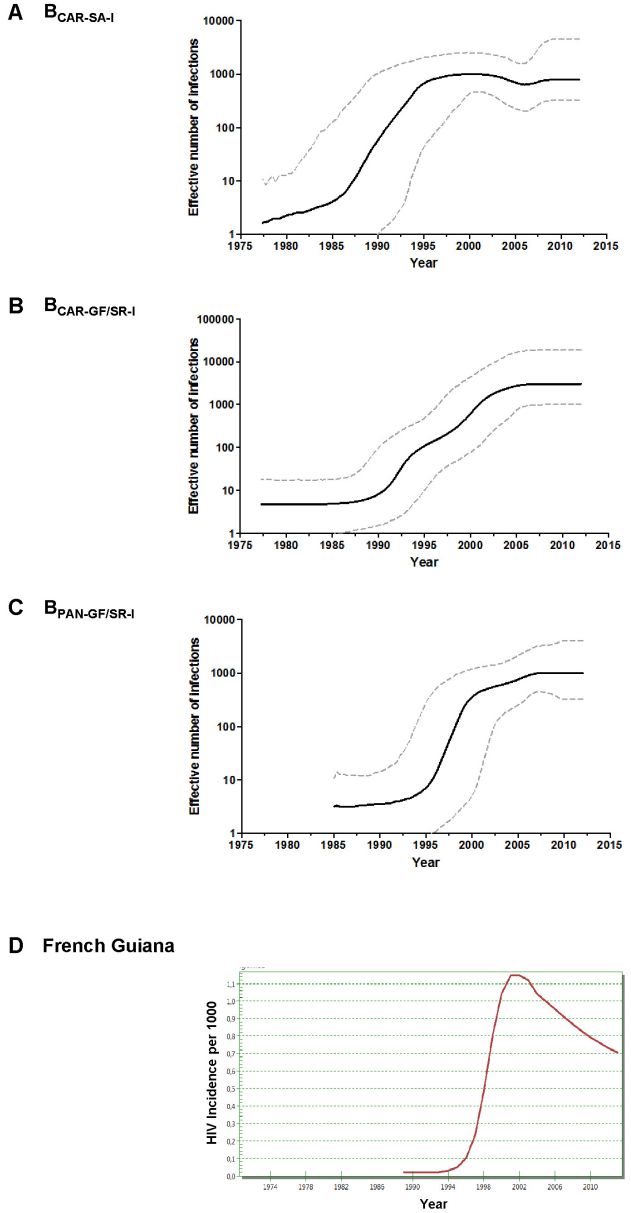


FIGURE 3 | (A–C) Demographic history of major HIV-1 B_{CAR} and $B_{PANDEMIC}$ lineages ($n > 30$ sequences) circulating in French Guiana and Suriname. Mean estimates of the effective number of infections (*Ne*) (solid line) are shown together with the 95% HPD intervals (dashed curves) of the Bayesian skyline plot for lineages $B_{CAR-SA-I}$ **(A)**, $B_{CAR-GF/SR-I}$ **(B)**, and $B_{PAN-GF/SR-I}$ **(C)**. The vertical axes represent the estimated *Ne* on a logarithmic scale. Time scales are in calendar years. **(D)** Spectrum-estimated national HIV incidence temporal trend in French Guiana.

$B_{CAR-GF/SR-I}$ (5%), $B_{CAR-GF/SR-II}$ (0%), and $B_{CAR-GF/SR-III}$ (10%). Thus, Haitian migrants may provide an epidemiological link for sporadic B_{CAR} transmissions from Haiti into French Guiana but only a minor fraction of those migrants seems to be actively participating in sustaining the local B_{CAR} transmission networks.

Our phylogeographic analysis supports that major $B_{PANDEMIC}$ lineages most probably arose after dissemination of viral strains from North America into French Guiana ($B_{PAN-GF/SR-I}$) and Suriname ($B_{PAN-GF/SR-II}$) at around the middle 1980s and from South America into French Guiana ($B_{PAN-GF-I}$) at around the early 1990s. Despite the historical ties and intense human mobility between French Guiana/mainland France and Suriname/Netherlands (Cordova Alcaraz, 2012), our analyses only support a few sporadic disseminations of $B_{PANDEMIC}$ strains between these regions. This is consistent with very restricted HIV transmissions between Surinamese and Dutch individuals living in the Netherlands (Kramer et al., 2011) and with the extremely low prevalence of B_{CAR} strains in mainland France and the Netherlands (Cabello et al., 2016). Similarly, no intense $B_{PANDEMIC}$ fluxes were detected between French Guiana and Northern Brazil despite the intense human mobility and the favorable social conditions for the spread of HIV in the border region between Amapá (Northern Brazil) and French Guiana (Bourdier, 2005; Soares et al., 2011; Parriault et al., 2015; Collomb and Fouck, 2016).

Although the high population mobility promote a frequent HIV flux between French Guiana and Suriname, our results support some level of geographic subdivision for some major subtype B lineages. We found that lineages $B_{CAR-SA-I}$ and $B_{PAN-GF-I}$ comprise a much larger proportion of sequences from French Guiana than from Suriname, have their origin traced to French Guiana and mostly comprise Guianese subjects followed-up at the capital city, Cayenne. Lineages $B_{CAR-GF/SR-I}$ and $B_{PAN-GF/SR-II}$, by contrast, were much more prevalent in Suriname than in French Guiana, had their origin traced to Suriname and mostly comprise subjects followed-up at Saint Laurent du Maroni, a French Guianese city located at the border with Suriname that attend a large number of patients from that country (Nacher et al., 2010; Jaries et al., 2017). We also detected a larger proportion of non-clustered B_{CAR} and $B_{PANDEMIC}$ infections in French Guiana relative to Suriname, perhaps reflecting the greater influence of migrations and/or the larger number of sequences analyzed from French Guiana.

The study of the epidemiological characteristics of HIV-infected subjects from French Guiana revealed that B_{CAR} and $B_{PANDEMIC}$ viruses have been disseminated between both MSM and heterosexual individuals. The MSM group has a much greater proportion of $B_{PANDEMIC}$ viruses (71%) than the heterosexual one (34%), but the role of MSM individuals in local spread of $B_{PANDEMIC}$ viruses greatly varied across transmission chains. While an important fraction of subjects infected with lineages $B_{PAN-GF/SR-II}$ (14%) and $B_{PAN-GF-I}$ (27%) were MSM; no MSM individuals were detected within the largest local $B_{PAN-GF/SR-I}$ transmission network. Overall, a large proportion of HIV-1 infections among heterosexuals (57%) and MSM (50%) in French Guiana can be ascribed to the seven major subtype B transmission chains here detected. More importantly, the proportion of individuals within these large cluster networks increased from 40% among those with HIV diagnosis between 1990 and 2005 to 62% among those diagnosed between 2006 and 2012. These data clearly support that large transmission clusters are a major driving force sustaining the recent dissemination of B_{CAR} and $B_{PANDEMIC}$ epidemics in French Guiana. These results

also emphasize that early detection and treatment as well as Pre Exposure Prophylaxis targeting people being part of large transmission chains should have a significant impact on reducing the spread of HIV-1 in French Guiana.

The analysis of clinical characteristics of individuals from French Guiana revealed a significant trend for higher RNA viral loads among $B_{PANDEMIC}$-infected relative to B_{CAR}-infected subjects, despite no significant differences in clinical condition or $CD4^+$ T cell counts between groups. Previous studies suggested that viral genetic factors contribute to the heritability and variation of viral load set point (Bonhoeffer et al., 2015; Bertels et al., 2018) and may also modulate viral replication (Ronsard et al., 2017a,b). Whether differences in viral load here observed between subjects reflect selective advantages for BPANDEMIC strains over B_{CAR} ones cannot be answered by the present study. Future *in vitro* studies comparing the replication dynamics of B_{CAR} and $B_{PANDEMIC}$ strains and studies of carefully controlled retrospective cohorts could further explore this possibility. Characterization of full-length viral genomes should also be done in future studies to detect potential intra-subtype B_{CAR}/$B_{PANDEMIC}$ recombinant lineages.

Although major lineages $B_{PAN\text{-}GF/SR\text{-}I}$, $B_{CAR\text{-}SA\text{-}I}$, and $B_{CAR\text{-}GF/SR\text{-}I}$ showed significant differences regarding city of residence, country of origin and RNA viral load, they exhibited roughly comparable mean epidemic growth rates (0.30–0.46 $years^{-1}$). Transactional sex and concurrent sexual partnerships were pointed as important drivers of the HIV epidemic in French Guiana (Nacher et al., 2010) and most subjects (>90%) infected by major Guianese/Surinamese lineages were heterosexuals. Interestingly, the epidemic growth rates here obtained were similar to those estimated for B_{CAR} and $B_{PANDEMIC}$ lineages spreading in countries from the Caribbean and Central America with generalized heterosexual epidemics (0.35–0.45 $years^{-1}$) (Cabello et al., 2014; Mendoza et al., 2014); but lower than those estimated for $B_{PANDEMIC}$ lineages mainly transmitted among MSM networks (0.75–1.55 $years^{-1}$) (Hue et al., 2005; Zehender et al., 2010; Chen et al., 2011; Delatorre and Bello, 2013; Worobey et al., 2016). This supports that the characteristics of the existing transmission network is a major driving force of the epidemic potential of different HIV-1 subtype B lineages.

In summary, this study highlights that HIV epidemics in French Guiana and Suriname are highly intermixed and that about 60% of HIV-1 subtype B infections in those countries resulted from the expansion of multiple B_{CAR} and $B_{PANDEMIC}$ viral strains probably introduced between the middle 1970s and the early 1990s. Major B_{CAR} and $B_{PANDEMIC}$ local lineages have independently spread among males and females of different age and risk groups from both countries and substantially contribute to the ongoing epidemic. Some associations between the infecting B_{CAR}/$B_{PANDEMIC}$ lineage and epidemiological (geographic location and country of birth) and clinical (RNA viral load) variables were detected among individuals sampled in French Guiana, but no major differences in the epidemic potential of major B_{CAR} and $B_{PANDEMIC}$ lineages were observed.

AUTHOR CONTRIBUTIONS

GB and VL conceived and designed the study. MN, ED, and VL collected and analyzed the epidemiological data. MN performed the estimation of the HIV incidence temporal trend in French Guiana. FD performed the phylogenetic analyses. DM and GB performed the phylodynamics inferences and produced all figures. GB, MN, and VL wrote the manuscript. All authors discussed and reviewed the manuscript.

FUNDING

This study has benefited from a European Commission "REGPOT-CT-2011-285837-STRonGer" grant within the FP7 and an "Investissement d'Avenir" grant managed by the Agence Nationale de la Recherche (CEBA, Ref. ANR-10-LABX-25-01). DM was funded by fellowships from "Agencia Nacional de Investigación e Innovación (ANII-Uruguay)" and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil)".

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01738/full#supplementary-material

FIGURE S1 | Lineage assignment of HIV-1 subtype B *pol* sequences from French Guiana and Suriname. **(A,C)** ML phylogenetic trees of HIV-1 subtype B *pol* PR/RT sequences (~1,000 nt) from French Guiana (n = 271) and Suriname (n = 90) together with representative sequences of the $B_{PANDEMIC}$ (US = 165 and France = 135) and the B_{CAR} (Caribbean = 200) lineages. Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). Asterisks point to highly supported (SH-*aLRT* > 0.90) key nodes of subtype B and the $B_{PANDEMIC}$ lineage. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. **(B,D)** Estimated proportion of B_{CAR} and $B_{PANDEMIC}$ lineages among HIV-1 subtype B infected individuals from French Guiana and Suriname according to the ML analyses. ND, not determined.

FIGURE S2 | Analysis of HIV-1 subtype B *pol* sequences with no lineage assignment. Similarity plots generated using SimPlot are shown on the left. Plots displays the percent genetic similarity (*y* axis) of each subtype B query sequence from French Guiana/Suriname to the reference B_{CAR} (blue line), $B_{PANDEMIC}$ (red line), and subtype D (black line) strains along the *pol* gene fragment (*x* axis; nucleotides 2,253–3,275 of reference strain HXB2). The analysis was performed with a window size of 250 nucleotides (nt) and a step size of 10 nt. Dashed vertical lines delimitate the *pol* gene fragments were most query sequences displayed highest similarity to B_{CAR} references (fragment I), near equal similarity to both B_{CAR} and $B_{PANDEMIC}$ references (fragment II), and highest similarity to $B_{PANDEMIC}$ references (fragment III). The GenBank accession number of each sequence is indicated in the upper left. NJ phylogenetic trees of subtype B query sequences from French Guiana/Suriname combined with B_{CAR} (blue circles), $B_{PANDEMIC}$ (red circles), and subtype D (black circles) strains at *pol* fragments I and III are shown on the left. The subtype/lineage assignment of each query sequence at each fragment is indicated in the upper middle. Subtype D sequences were used as outgroups and the branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. Low bootstrap support values were obtained for both subtype B (<70%) and $B_{PANDEMIC}$ (<30%) monophyletic groups in all phylogenetic trees.

FIGURE S3 | Map summarizing viral migration events of major HIV-1 B_{CAR} lineages circulating in French Guiana and Suriname. Arrows between locations represent branches in the Bayesian phylogenetic tree along, which location's transitions occur. Map was created with Quantum GIS (QGIS) software from templates obtained from d-maps.com (Caribbean: http://d-maps.com/pays.php?num_pay=118&lang=es; South America: http://d-maps.com/carte.php?num_car=28522&lang=en).

TABLE S1 | HIV-1 subtype B pol (PR/RT) reference sequences of BPANDEMIC and BCAR clades used for subtype B clade assignment.

TABLE S2 | HIV-1 subtype B pol (PR/RT) reference sequences of BPANDEMIC and BCAR clades used for identification of Guianese/Surinamese subtype B clades.

TABLE S3 | HIV-1 BCAR pol (PR/RT) sequences used for Bayesian phylogeographic analysis.

TABLE S4 | HIV-1 BPANDEMIC pol (PR/RT) sequences used for Bayesian phylogeographic analysis.

TABLE S5 | Clade assignment of HIV-1 subtype B subtype pol sequences from French Guiana and Suriname.

TABLE S6 | Phylogenetic clustering of HIV-1 BCAR and BPANDEMIC pol sequences from French Guiana and Suriname.

TABLE S7 | Prevalence of major HIV-1 BCAR lineages circulating in French Guiana and Suriname.

TABLE S8 | Prevalence of major HIV-1 BPANDEMIC lineages circulating in French Guiana and Suriname.

TABLE S9 | Best fit demographic models for major HIV-1 subtype B lineages circulating in French Guiana and Suriname.

TABLE S10 | Epidemiological information of subjects from French Guiana infected by major HIV-1 BCAR and BPANDEMIC lineages.

REFERENCES

Abdoel Wahid, F., Sno, R., Darcissac, E., Lavergne, A., Adhin, M. R., and Lacoste, V. (2016). HIV-1 genetic diversity and drug resistance mutations among treatment-naive adult patients in suriname. *AIDS Res. Hum. Retroviruses* 32, 1223–1228. doi: 10.1089/AID.2016.0161

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., and Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* 29, 2157–2167. doi: 10.1093/molbev/mss084

Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Fellay, J., Gunthard, H. F., et al. (2018). Dissecting HIV virulence: heritability of setpoint viral load, CD4+ T-cell decline, and per-parasite pathogenicity. *Mol. Biol. Evol.* 35, 27–37. doi: 10.1093/molbev/msx246

Bonhoeffer, S., Fraser, C., and Leventhal, G. E. (2015). High heritability is compatible with the broad distribution of set point viral load in HIV carriers. *PLoS Pathog.* 11:e1004634. doi: 10.1371/journal.ppat.1004634

Borland, R., Faas, L., Marshall, D., Mclean, R., Schroen, M., Smit, M., et al. (2004). *HIV/AIDS and Mobile Populations in the Caribbean: A Baseline Assessment. International Organization for Migration (IOM)*. Available at: https://publications.iom.int/books/hivaids-and-mobile-populations-caribbean-baseline-assessment

Bourdier, F. (2005). L'avancée du sida dans les zones frontalières guyano-brésiliennes [in French]. *Hommes Migr.* 1256, 116–129. doi: 10.3406/homig.2005.4376

Cabello, M., Junqueira, D. M., and Bello, G. (2015). Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. *AIDS* 29, 483–492. doi: 10.1097/QAD.0000000000000552

Cabello, M., Mendoza, Y., and Bello, G. (2014). Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. *PLoS One* 9:e106045. doi: 10.1371/journal.pone.0106045

Cabello, M., Romero, H., and Bello, G. (2016). Multiple introductions and onward transmission of non-pandemic HIV-1 subtype B strains in North America and Europe. *Sci. Rep.* 6:33971. doi: 10.1038/srep33971

Chen, J. H., Wong, K. H., Chan, K. C., To, S. W., Chen, Z., and Yam, W. C. (2011). Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS One* 6:e25286. doi: 10.1371/journal.pone.0025286

Collomb, G., and Fouck, S. M. L. (2016). *Mobilités, Ethnicités, Diversité Culturelle: La Guyane Entre Surinam et Brésil Eléments de Compréhension de la Situation Guyanaise [in French]*. Matoury: Ibis Rouge Editions.

Corbin, H. P. (2007). *Brazilian Migration to Guyana as a Livelihood Strategy: A Case Study Approach*. Available at: http://www.repositorio.ufpa.br/jspui/bitstream/2011/1966/1/Dissertacao_BrazilianMigrationGuyana.pdf

Cordova Alcaraz, R. (2012). *Migratory Routes and Dynamics Between Latin American and Caribbean (LAC) Countries and Between LAC and the European Union. International Organization for Migration (IOM)*. Available at: https://publications.iom.int/system/files/pdf/migration_routes_digital.pdf

Darcissac, E., Nacher, M., Adriouch, L., Berlioz-Arthaud, A., Boukhari, R., Couppie, P., et al. (2016). HIV-1 Pol gene polymorphism and antiretroviral resistance mutations in treatment-naive adult patients in French Guiana between 2006 and 2012. *AIDS Res. Hum. Retroviruses* 32, 801–811. doi: 10.1089/AID.2016.0048

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., et al. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21, 3797–3800. doi: 10.1093/bioinformatics/bti607

Delatorre, E., and Bello, G. (2013). Phylodynamics of the HIV-1 epidemic in Cuba. *PLoS One* 8:e72448. doi: 10.1371/journal.pone.0072448

Divino, F., de Lima Guerra Corado, A., Gomes Naveca, F., Stefani, M. M., and Bello, G. (2016). High prevalence and onward transmission of non-pandemic HIV-1 subtype B clades in Northern and Northeastern Brazilian regions. *PLoS One* 11:e0162112. doi: 10.1371/journal.pone.0162112

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

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307–1320.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103

Ferreira, M. A. R., and Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* 36, 355–368. doi: 10.1002/cjs.5550360302

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and 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

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–W559. doi: 10.1093/nar/gki352

Hue, S., Pillay, D., Clewley, J. P., and Pybus, O. G. (2005). Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4425–4429. doi: 10.1073/pnas.0407534102

Hyles, J. R. (2014). *Guiana and the Shadows of Empire: Colonial and Cultural Negotiations at the Edge of the World*. Lanham, MD: Lexington Books.

Jaries, R., Vantilcke, V., Clevenbergh, P., Adoissi, J., Boukhari, R., Misslin, C., et al. (2017). Population movements and the HIV cascade in recently diagnosed

patients at the French Guiana -Suriname border. *AIDS Care* 29, 1448–1452. doi: 10.1080/09540121.2017.1291899

Junqueira, D. M., De Medeiros, R. M., Graf, T., and Almeida, S. E. (2016). Short-term dynamic and local epidemiological trends in the South American HIV-1B epidemic. *PLoS One* 11:e0156712. doi: 10.1371/journal.pone.0156712

Kazanji, M., Lavergne, A., Pouliquen, J. F., Magnien, C., Bissuel, F., Marty, C., et al. (2001). Genetic diversity and phylogenetic analysis of human immunodeficiency virus type 1 subtypes circulating in French Guiana. *AIDS Res. Hum. Retroviruses* 17, 857–861. doi: 10.1089/08892220175025 2052

Kramer, M. A., Cornelissen, M., Paraskevis, D., Prins, M., Coutinho, R. A., Van Sighem, A. I., et al. (2011). HIV transmission patterns among The Netherlands, Suriname, and The Netherlands Antilles: a molecular epidemiological study. *AIDS Res. Hum. Retroviruses* 27, 123–130. doi: 10.1089/aid.2010. 0115

Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/ journal.pcbi.1000520

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Magiorkinis, G., Angelis, K., Mamais, I., Katzourakis, A., Hatzakis, A., Albert, J., et al. (2016). The global spread of HIV-1 subtype B epidemic. *Infect. Genet. Evol.* 46, 169–179. doi: 10.1016/j.meegid.2016.05.041

Mendoza, Y., Martinez, A. A., Castillo Mewa, J., Gonzalez, C., Garcia-Morales, C., Avila-Rios, S., et al. (2014). Human immunodeficiency virus type 1 (HIV-1) subtype B epidemic in panama is mainly driven by dissemination of country-specific clades. *PLoS One* 9:e95360. doi: 10.1371/journal.pone.0095360

Mir, D., Cabello, M., Romero, H., and Bello, G. (2015). Phylodynamics of major HIV-1 subtype B pandemic clades circulating in Latin America. *AIDS* 29, 1863–1869. doi: 10.1097/QAD.0000000000000770

Nacher, M., Vantilcke, V., Parriault, M. C., Van Melle, A., Hanf, M., Labadie, G., et al. (2010). What is driving the HIV epidemic in French Guiana? *Int. J. STD AIDS* 21, 359–361. doi: 10.1258/ijsa.2010.009570

Paraskevis, D., Pybus, O., Magiorkinis, G., Hatzakis, A., Wensing, A. M., Van De Vijver, D. A., et al. (2009). Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology* 6:49. doi: 10.1186/1742-4690-6-49

Parriault, M. C., Van Melle, A., Basurko, C., Gaubert-Marechal, E., Macena, R. H., Rogier, S., et al. (2015). HIV-testing among female sex workers on the border between Brazil and French Guiana: the need for targeted interventions. *Cad. Saude Publica* 31, 1615–1622. doi: 10.1590/0102-311X00138514

Pereira, M. C. (2006). Processos migratoìrios na fronteira Brasil-Guiana. *Estud. Avançados* 20, 209–219. doi: 10.1590/S0103-40142006000200016

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Rambaut, A. (2009). *FigTree v1.4: Tree Figure Drawing Tool*. Available at: http: //tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., and Drummond, A. (2007). *Tracer v1.6*. Available at: http://tree.bio. ed.ac.uk/software/tracer/

Rambaut, A., Lam, T. T., Carvalho, L. M., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Ronsard, L., Ganguli, N., Singh, V. K., Mohankumar, K., Rai, T., Sridharan, S., et al. (2017a). Impact of genetic variations in HIV-1 Tat on LTR-mediated transcription via TAR RNA interaction. *Front. Microbiol.* 8:706. doi: 10.3389/ fmicb.2017.00706

Ronsard, L., Rai, T., Rai, D., Ramachandran, V. G., and Banerjea, A. C. (2017b). In silico analyses of subtype specific HIV-1 Tat-TAR RNA interaction reveals the structural determinants for viral activity. *Front. Microbiol.* 8:1467. doi: 10.3389/fmicb.2017.01467

Soares, C. L., De Souza Oliveira, B., and De Souza Pinto, M. J. (2011). Trabalhadores brasileiros na Guiana Francesa: entre a invisibilidade e o desemprego. *PRACS Rev. Hum. Curso Ciênc. Soc.* 4, 129–142.

Suchard, M. A., and Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25, 1370–1376. doi: 10.1093/bioinformatics/ btp244

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876

UNAIDS (2013). *Report on the Global AIDS Epidemic*. Available at: http: //www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/ 2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

Worobey, M., Watts, T. D., Mckay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., et al. (2016). 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539, 98–101. doi: 10.1038/ nature19827

Zehender, G., Ebranati, E., Lai, A., Santoro, M. M., Alteri, C., Giuliani, M., et al. (2010). Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J. Acquir. Immune Defic. Syndr.* 55, 156–160. doi: 10.1097/QAI.0b013e3181eb3002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 September 2018
doi: 10.3389/fmicb.2018.02150

Universal Target Capture of HIV Sequences From NGS Libraries

*Julie Yamaguchi[1], Ana Olivo[1], Oliver Laeyendecker[2], Kenn Forberg[1], Nicaise Ndembi[3], Dora Mbanya[4,5], Lazare Kaptue[6], Thomas C. Quinn[2], Gavin A. Cloherty[1], Mary A. Rodgers[1] and Michael G. Berg[1]**

[1] Infectious Diseases Research, Abbott Diagnostics, Chicago, IL, United States, [2] National Institute of Allergy and Infectious Diseases, NIH, Baltimore, MD, United States, [3] Institute of Human Virology, Abuja, Nigeria, [4] Université de Yaoundé 1, Yaoundé, Cameroon, [5] University of Bamenda, Bamenda, Cameroon, [6] Université des Montagnes, Bangangté, Cameroon

Background: Global surveillance of viral sequence diversity is needed to keep pace with the constant evolution of HIV. Recent next generation sequencing (NGS) methods have realized the goal of sequencing circulating virus directly from patient specimens. Yet, a simple, universal approach that maximizes sensitivity and sequencing capacity remains elusive. Here we present a novel HIV enrichment strategy to yield near complete genomes from low viral load specimens.

Methodology: A non-redundant biotin-labeled probe set (HIV-xGen; $n = 652$) was synthesized to tile all HIV-1 (groups M, N, O, and P) and HIV-2 (A and B) strains. Illumina Nextera barcoded libraries of either gene-specific or randomly primed cDNA derived from infected plasma were hybridized to probes in a single pool and unbound sequences were washed away. Captured viral cDNA was amplified by Illumina adaptor primers, sequenced on a MiSeq, and NGS reads were demultiplexed for alignment with CLC Bio software.

Results: HIV-xGen probes selectively captured and amplified reads spanning the entirety of the HIV phylogenetic tree. HIV sequences clearly present in unenriched libraries of specimens but previously not observed due to high host background levels, insufficient sequencing depth or the extent of multiplexing, were now enriched by >1,000-fold. Thus, xGen selection not only substantially increased the depth of existing sequence, but also extended overall genome coverage by an average of 40%. We characterized 50 new, diverse HIV strains from clinical specimens and demonstrated a viral load cutoff of approximately log 3.5 copies/ml for full length coverage. Genome coverage was <20% for 5/10 samples with viral loads <log 3.5 copies/ml and >90% for 35/40 samples with higher viral loads.

Conclusions: Characterization of >20 complete genomes at a time is now possible from a single probe hybridization and MiSeq run. With the versatility to capture all HIV strains and the sensitivity to detect low titer specimens, HIV-xGen will serve as an important tool for monitoring HIV sequence diversity.

Keywords: next-generation sequencing, HIV, HIV diversity, target enrichment, xGen, virus surveillance

OPEN ACCESS

Edited by:
Michael M. Thomson, Instituto de Salud Carlos III, Spain

Reviewed by:
Hirotaka Ode, Nagoya Medical Center (NHO), Japan
Yorifumi Satou, Kumamoto University, Japan

***Correspondence:**
Michael G. Berg
michael.berg@abbott.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *05 June 2018*
Accepted: *22 August 2018*
Published: *13 September 2018*

Citation:
Yamaguchi J, Olivo A, Laeyendecker O, Forberg K, Ndembi N, Mbanya D, Kaptue L, Quinn TC, Cloherty GA, Rodgers MA and Berg MG (2018) Universal Target Capture of HIV Sequences From NGS Libraries. Front. Microbiol. 9:2150. doi: 10.3389/fmicb.2018.02150

INTRODUCTION

As the HIV epidemic continues, surveillance of HIV diversity is essential to monitor the emergence of new subtypes and the presence of new strains (Hemelaar, 2013). Knowing which subtypes or recombinants historically predominate in a geographic region and whether the identity and proportion of new infections reflect a static or changing landscape will help inform appropriate interventions. For example, in Europe where the majority were once subtype B, 50% of new infections are non-B or recombinants (Semaille et al., 2007). The number of non-B infections in the US has also been on the rise (Pyne et al., 2013; Oster et al., 2017). Of equal concern is knowing whether mutations have arisen in current strains that could lead to misdiagnosis of HIV status, under-quantification of patients on antiretroviral therapy, or inadequate screening of the blood supply (Brennan et al., 2006).

Next generation sequencing (NGS) has been applied to the fight against HIV in several capacities, including monitoring for drug resistance, cell tropism determination, superinfection studies, network analysis and transmission patterns, and intra-patient quasi-species detection (Bimber et al., 2010; Archer et al., 2012; Redd et al., 2012; Swenson et al., 2012; Giallonardo et al., 2014). These typically involve high-throughput amplicon sequencing of defined sub-genomic regions such as *pol* or *env*. By contrast, metagenomic approaches using random priming permit an assessment of the full extent of sequence diversity in strains and can be applied to surveillance. Complete genomes of HIV and co-infecting viruses can be obtained directly from patient specimens, provided viral loads are high enough (Luk et al., 2015; Yamaguchi et al., 2017). To increase sensitivity, the HIV-SMART method utilizes reverse primers in conserved sequences of HIV fused to a tag to facilitate combined virus-specific priming and amplification (Berg et al., 2016; Rodgers et al., 2017a). In an alternate HIV-specific approach, individual long fragments (e.g., >2 kb) amplified with HIV primer pairs can be converted to NGS libraries to achieve detection limits nearing that of PCR, although the process is more labor intensive (Gall et al., 2012). What remains elusive is a simple, universal NGS approach for surveillance that balances sensitivity and cost, allowing numerous specimens of any titer and from any HIV group to be sequenced simultaneously while also maximizing NGS capacity and minimizing the use of resources.

To increase the sensitivity of microbial metagenomics from patient specimens, numerous pre-treatment approaches have been applied to reduce host background, including nuclease pre-treatment, rRNA depletion, filtration, centrifugation, etc. (Hall et al., 2014; Conceição-Neto et al., 2015). An alternate means of enriching for viral reads after NGS library preparation has shown great promise. Briese, et al synthesized nearly 2 million biotin-labeled probes to cover all coding regions of vertebrate viruses, hybridized them to conventional high throughput sequencing libraries containing spiked in viral nucleic acid, and amplified sequences captured on streptavidin beads. The VirCapSeq-VERT method yielded a 100–10,000-fold increase in viral reads from patient specimens (Briese et al., 2015). An analogous approach tailored specifically for hepatitis C virus (HCV) has been equally successful, tiling all seven genotypes with just 1,100 probes (Bonsall et al., 2015). HIV probes tiling only subtype B showed in principle that genomes of integrated provirus can be selectively amplified from host DNA (Miyazato et al., 2016). As with HCV, HIV-1 and HIV-2 exhibit substantial diversity and therefore probes inclusive of all subtypes and groups are needed for surveillance. Here we developed a comprehensive probe set to accommodate the full spectrum of HIV strains encountered. We demonstrate that the HIV-xGen method is highly effective, particularly for low titer patient specimens, moving the field closer toward a universal, sensitive, high volume NGS solution.

MATERIALS AND METHODS

Specimens

Specimens were collected by the Abbott Global Surveillance Program through collaborations in Cameroon, Uganda, South Africa, Senegal, and Thailand. All specimens were de-identified and obtained according to local regulations in each country at the time of collection between 1998 and 2016, including local IRB approval when required. Specimens were identified as HIV-1 positive by rapid diagnostic tests done in source countries before sequence analysis and testing at Abbott Laboratories (Swanson et al., 2000; Brennan et al., 2008; Rodgers et al., 2017b). De-identified specimens from the Democratic Republic of Congo were obtained in 1987 as part of Project SIDA by the US National Institute of Allergy and Infectious Diseases (Cohen, 1997).

Specimen Pretreatment and Extraction

Plasma specimens were pre-treated with Ultra-pure benzonase (Sigma, St. Louis MO) for 3 h at 37°C and extracted on the *m*2000*sp* (Abbott Laboratories, Des Plaines IL) with the RNA protocol (500 µl input/70 µl elute) as described (Berg et al., 2016).

HIV Viral Loads

Viral loads were determined by the HIV-1 RealTi*m*e (Abbott Laboratories, Des Plaines IL) assay as described (Berg et al., 2016).

cDNA Synthesis and Nextera XT Library Production

Viral RNA was concentrated to 10 µl with RNA Clean and Concentrator-5 spin columns (Zymo Research, CA) as described (Rodgers et al., 2017a). For gene-specific reverse transcription, cDNA was generated with HIV-SMART essentially as described (Berg et al., 2016; Rodgers et al., 2017a). Minor modifications described here were made to primer concentrations, cDNA input and PCR cycling conditions to reduce loss and boost sensitivity. The pool of six HIV-SMART primers used was 300 nM final for each (formerly 1 µM), the entire 10 µl of SMART cDNA (without dilution) was used as input for SMART amplification, and 18 rounds (formerly 35) of PCR were performed. HIV-SMART libraries were then purified with 1.4X vol of AMPpure beads, eluted in 12 µl of EB buffer, and 5 µl (undiluted) was used as template for Nextera XT reactions. For reverse transcription by random primers, metagenomic libraries were prepared by using Superscript III (SSRTIII) 1st Strand reagents

(Life Technologies) followed by 2nd strand synthesis with Sequenase V2.0 T7 DNA pol (Affymetrix). Double stranded cDNA was recovered with DNA clean and concentrator spin columns (Zymo Research) and eluted in 7 μl. HIV-SMART and SSRTIII libraries were tagmented by Nextera XT and amplified for 16 and 24 cycles, respectively, with Set A indices lacking 5′ biotin tags (see below), according to manufacturer instructions (Illumina, Carlsbad CA). Nextera libraries were purified with Agencourt AMPpure XP beads (Beckman Coulter), visualized on a BioAnalyzer TapeStation (Agilent), and quantified on a Qubit instrument (Life Technologies) with dsDNA broad range reagents (Molecular Diagnostics).

Design of HIV xGen Probe Set

For HIV-1 group M, approximately 10 complete reference sequences from each subtype (A-K, plus CRF02) and representing a diversity of countries of origin were selected from the Los Alamos National Lab 2015 alignment. HIV-xGen probes ($n = 82$) were first designed against the consensus sequence derived from this reduced alignment. All degenerate base codes (e.g., R, Y) were replaced with specific, majority base call designations to permit probe synthesis. Next, the consensus was compared to individual subtypes by scanning each sequence in 100 bp windows to identify regions with <80% identity. Where diversity exceeded this cutoff and substitutions were found to be common across several subtypes, additional probes ($n = 101$) tiling 122–366 bp segments were designed to ensure complete capture of these regions. In order to adequately tile stretches with >10% diversity (e.g., *env*), probes were either designed against the entire individual sequence or to those sequences flanking the variable region. Consensus sequences were similarly obtained to design probe sets for HIV-1 groups N ($N = 78$), O ($N = 83$) and P ($N = 83$), and HIV-2 groups A ($N = 87$) and B ($N = 86$), each supplemented with additional probes ($N = 11$, 31, 6, 3, and 0, respectively) covering regions of increased diversity.

Synthesis of HIV xGen Probes, Nextera Barcoding Primers, and Blocking Oligos

120 nt probes encoding the sense strand of HIV with 1 nt overlap and modified with a 5′ biotin tag were synthesized at Integrated DNA Technologies (IDT, Coralville IA). Mini-pools made at 3 pmol/probe were resuspended in 15 μl of TE to a storage concentration of 0.2 pmol/probe/μl.

Custom DNA Ultramers (4 nmol) of Nextera XT Set A indices (N701-N715 and S502-S511) lacking a biotin label and containing two 3′ phosphothioate modifications were synthesized at IDT and working stocks were each diluted to 5 μM. Blocking oligos complementary to Nextera Set A i5 and i7 index primers (e.g., P5/P7 adaptors_8 nt barcodes_R1/R2 sequencing primers) were also synthesized at IDT at a 1 μl/reaction working concentration.

xGen Hybridization, Washes and Library Amplification

The enrichment for HIV reads was performed essentially as described in the *Hybridization capture of DNA libraries using xGen® Lockdown® probes and reagents* protocol from IDT. Individually barcoded Nextera libraries ($n = 6$–26) were pooled to have a minimum 500 ng of cDNA and then combined with 5 μg of Cot-1 DNA carrier and 1 μl of each blocking oligo. Samples were dried for >15 min in a SpeedVac set at 45°C. Pelleted material was resuspended in 8.5 μl of 2X hybridization buffer, 2.7 μl of Hybridization Buffer Enhancer, and 1.8 μl of nuclease-free water. After a 10 min denaturation at 95°C, 4 μl of a working probe stock (100 attomoles/probe/μl) was added and mixed to bring the final volume to 17 μl. Hybridizations were incubated for 4 h at 65°C. M-270 Streptavidin beads (100 μl per capture) were equilibrated in Bead Wash buffer, pelleted by magnet, and mixed with hybridizations for another 45-min incubation at 65°C, vortexing every 12 min to ensure beads remained in solution. Washes were performed as recommended and beads were resuspended in 20 μl of nuclease-free water for an initial 12 cycles of amplification with the KAPA PCR mix. Agencourt AMPpure beads (1.5X volume/75 μl) were added to PCR reactions and captured/washed material was eluted in 20 μl. A repeat KAPA amplification of 10 cycles followed by AMP-Pure clean-up was performed and libraries were visualized on an Agilent 2200 TapeStation and quantified with a Qubit fluorometer using the dsDNA high-sensitivity kit.

NGS Analysis

Analysis was performed as described (Berg et al., 2016; Rodgers et al., 2017a). Barcodes were parsed on the MiSeq instrument and filtered for Q-scores above 30. Fastq files were imported into CLC Genomics Workbench 9.0 software (CLC bio/Qiagen, Aarhus Denmark, version 9.0) where reads below 70 nt were discarded and the SMART adaptor was removed separately from both strands. When no preliminary Sanger sequence was available, raw data was simultaneously mapped to multiple HIV reference sequences to determine the subtype/group with the greatest identity. The following alignment settings were applied: match = 1, mismatch = 2, insertion = 3, deletion = 3, length fraction = 0.7, and similarity fraction = 0.8. An iterative approach was used to derived the final sequence, whereby the initial consensus served as the reference to refine the consensus in a second round of alignment. The raw NGS data (see below) was realigned to the final consensus sequence to generate NGS run statistics found in **Tables 1**, **2**.

Removal of Contaminating Sequences

To detect possible contaminating reads originating from a different set of barcodes, raw data from one sample was individually mapped to the consensus sequences of samples sequenced in the same run, requiring a similarity fraction of 0.99 (e.g., ≥99% identical). Mapped reads were removed from the fasta file. Unmapped reads (e.g., unique to the sample of interest) were collected and realigned to an appropriate reference to generate a consensus.

Sequence Gap Closure

RT-PCR and Sanger (population) sequencing were performed as described to fill gaps in NGS sequence data (Rodgers et al., 2017a). Primer pairs used and their sequences are listed in the

TABLE 1 | NGS data for samples sequenced ± enrichment with HIV xGen.

Specimen ID	Country	Group or M subtype	Viral Load	Input copy no.	Library type	Total reads	HIV reads	% HIV reads	Genome coverage%	Avg. coverage depth	Std. Dev.
O-LA34	Cameroon	HIV-1 O	8.06	8019	*HIV-SMART	1,336,488	50,865	3.80	100	986	589
–	–	–	–	20667	SSIII-xGen	1,477,954	1,436,992	97.23	100	14,576	11,026
N-LA28	Cameroon	HIV-1 N	7.61	nt	*HIV-SMART	1,564,286	155,741	10.00	100	3,422	1,604
–	–	–	–	7333	SSIII-xGen	2,033,250	2,020,867	99.39	100	22,538	17,020
2A-LA38	Senegal	HIV-2A	6.95		*Ovation single cell	3,029,490	40,402	1.33	100	573.43	445.55
–	–	–	–	28355	SSIII	37,384	152	0.41	72	2	2
–	–	–	–	–	SSIII-xGen	342,284	331,623	96.89	100	3,400	3,344
C-8119636	South Africa	HIV-1 C	5.62	8870	*HIV-SMART	5,446,060	367,224	6.74	99	6,481	4,483
–	–	–	–	85260	SSIII	586,936	15800	2.69	98	156	196
–	–	–	–	–	SSIII-xGen	400,458	393,350	98.23	100	4,042	5,648
02-421-10	Cameroon	HIV-1 CRF02	4.48	643	*HIV-SMART	1,418,404	42,153	2.97	61	1,201	1,127
–	–	–	–	6177	SSIII	412,134	80	0.02	37	1	1
–	–	–	–	–	SSIII-xGen	9,776	7,131	72.94	82	78	231
PBS1342†	DRC	HIV-1 URF	4.59	3830	HIV-SMART	662,360	3,511	0.53	86	19	94
–	–	–	–	–	HIV-SMART-xGen	691,384	669,074	96.77	95	8,673	14,978
–	–	–	–	–	SSIII	3,432,778	1,095	0.03	99	11	8
–	–	–	–	–	SSIII-xGen	786,548	747,840	95.07	97	8,746	13,219
PBS1191	DRC	HIV-1 G	3.86	1426	HIV-SMART	411,612	2,067	0.50	51	7	94
–	–	–	–	–	HIV-SMART-xGen	2,156,988	1,504,919	69.77	100	17,823	21,968
–	–	–	–	–	SSIII	1,276,656	306	0.02	45	2	4
–	–	–	–	–	SSIII-xGen	3,968,286	3,313,601	83.50	96	36,912	126,521
PBS888	DRC	HIV-1 A	3.47	581	HIV-SMART	1,037,832	208	0.02	56	2	4
–	–	–	–	–	HIV-SMART-xGen	2,172,982	1,880,055	86.52	100	25,171	53,285
–	–	–	–	–	SSIII	4,532,554	257	0.01	54	2	2
–	–	–	–	–	SSIII-xGen	1,699,092	925,860	54.49	96	10,689	23,552
70641	DRC	HIV-1 A1	5.78		HIV-SMART	1,361,655	51,646	3.79	99	947	1,385
	–	–	–	–	HIV-SMART-xGen	9,759,614	9,158,341	93.84	100	193,337	648,377
PBS5635†	DRC	HIV-1 D	5.31		HIV-SMART	1,255,489	12,653	1.01	95	259	320
	–	–	–	–	HIV-SMART-xGen	574,273	522,043	90.91	99	12,321	30,529
P406	DRC	HIV-1 G	5.21		HIV-SMART	2,206,004	4,676	0.20	98	76	98
	–	–	–	–	HIV-SMART-xGen	1,238,978	629,912	50.84	100	14,903	40,689
PBS1189‡	DRC	HIV-1 F1	4.98		HIV-SMART	1,978,086	2,260	0.11	82	41	76
	–	–	–	–	HIV-SMART-xGen	522,902	463,644	88.67	86	9,236	41,558
50†	DRC	HIV-1 A1 basal	4.96		HIV-SMART	1,662,501	1,731	0.10	83	29	38
	–	–	–	–	HIV-SMART-xGen	237,883	218,072	91.67	83	4,899	15,280
87-2580†	DRC	HIV-1 G	4.85		HIV-SMART	2,200,708	5,741	0.26	96	112	167
	–	–	–	–	HIV-SMART-xGen	529,218	221,335	41.82	96	5,088	10,026
PBS6126	DRC	HIV-1 A1	4.67		HIV-SMART	149,709	115	0.08	79	2.6	2.4
	–	–	–	–	HIV-SMART-xGen	28,606	11,644	40.70	100	4,034	276
2049‡	DRC	HIV-1 CRF37 basal	4.31		HIV-SMART	2,720,263	1,677	0.06	73	33	46
	–	–	–	–	HIV-SMART-xGen	10,689,440	9,321,568	87.20	90	201,793	340,069
PBS0724‡	DRC	HIV-1A1	4.02		HIV-SMART	107,044	1	0.00	2	0.02	0.16
	–	–	–	–	HIV-SMART-xGen	2,010,039	1,061,113	52.79	100	21,120	38,865
PBS1195	DRC	HIV-1 A2	3.77		HIV-SMART	605,867	51	0.01	54	1.2	1.4

(Continued)

TABLE 1 | Continued

Specimen ID	Country	Group or M subtype	Viral Load	Input copy no.	Library type	Total reads	HIV reads	% HIV reads	Genome coverage%	Avg. coverage depth	Std. Dev.
	–	–	–	–	HIV-SMART-xGen	85,628	45,233	52.83	100	1,039	1,847
2106	DRC	HIV-1 A1	3.76		HIV-SMART	1,664,855	598	0.04	80	13	14
	–	–	–	–	HIV-SMART-xGen	1,355,794	1,054,751	77.80	100	24,313	57,096
P4039	DRC	HIV-1 A1	3.58		HIV-SMART	419,163	132	0.03	71	2.8	3
	–	–	–	–	HIV-SMART-xGen	398,324	228,302	57.32	100	5,084	6,066
P3844	DRC	HIV-1 K	4.67		HIV-SMART	37,076	7	0.02	8	1.27	0.59
	–	–	–	–	HIV-SMART-xGen	1,883,032	992,447	52.7	97	23,300	57,444

*Previously published data reported in Berg et al. (2016).
†Genome completed (100%) with Sanger.
‡Genome coverage increased with Sanger but still incomplete.

Supplemental Information. Complete genomes were obtained for 144-26, 363-24, PBS1342, AB260, and 814-43 by combining Sanger with NGS data. Additional sequence covering gaps was obtained for 9505343, 129-26, 961-09, 577-27, 112-11, and 8128965.

Phylogenetic Analysis

Alignments, neighbor-joining trees, and recombination analysis by SIMPLOT were all performed as described (Berg et al., 2016).

Nucleotide Sequence Accession Numbers

Open reading frames for all 28 full and 3 near complete genomes were verified and annotated using SeqBuilder (DNASTAR Lasergene v11.2) and submitted to GenBank as project 2135293 under the following accessions: PBS888 (MH705133), PBS1191 (MH705134), PBS1342 (MH705135), 459-16 (MH705136), 5056135 (MH705137), 8128965 (MH705138), 9505343 (MH705139), 814-43 (MH705140), 363-24 (MH705141), 770-8 (MH705142), 42-877 (MH705143), A1699 (MH705144), 144-26 (MH705145), 112-11 (MH705146), AB260 (MH705147), 296 (MH705148), O-1225-51 (MH705149), 20-02 (MH705150), 70641 (MH705151), PBS5635 (MH705152), PBS6126 (MH705153), 2049 (MH705154), P406 (MH705155), P3844 (MH705156), P4039 (MH705157), 2106 (MH705158), PBS0724 (MH705159), PBS1189 (MH705160), 50 (MH705161), 87-2580 (MH705162), PBS1195 (MH705163). Raw data depleted of human sequences can be found in the SRA database under BioProject ID: PJRNA486839.

RESULTS

HIV-xGen Target Enrichment Strategy

To enable full genome characterization of all HIV strains, we designed xGen probes to selectively capture and amplify viral sequences from cDNA libraries. Individual alignments were compiled for HIV-1 group M (subtypes A-K and circulating recombinant form [CRF] 02), groups N, O and P, as well as HIV-2 groups A and B. A minimum 80% identity has been shown to be required for effective xGen probe hybridization of viral sequences (Bonsall et al., 2015). Therefore, to eliminate redundancy and the synthesis of a prohibitively expensive number of probes, a consensus sequence from each group was generated from which an initial set of 120 nt single-stranded DNA probes tiling the genome at 1X coverage was derived. Alignments were then scanned in 100 nt windows to identify regions of strong nucleotide conservation and those of considerable heterogeneity. The minimum number of probes were selected to tile the former (e.g., >80% identity in *pol*), whereas those for heterogeneous regions (e.g., <80% identity in *env*) found in individual subtypes and strains were added as needed. Using this approach, only 183 probes were required for HIV-1 group M, compared to several thousand that would have been needed if probes were designed against the entire genomic sequence of individual strains. A total of 651 probes covered all HIV-1 and HIV-2 strains (**Figure 1**).

cDNA from HIV-infected plasma can be synthesized by either random (Superscript RTIII; SSRTIII) or virus sequence-specific priming (HIV-SMART; **Figure 1**), followed by topoisomerase-mediated fragmentation, adaptor tagging and amplification with Nextera XT (Berg et al., 2015, 2016). Previously, despite a 17–20-fold increase in sensitivity over metagenomic (random primed) libraries, together with additional optimization of the HIV-SMART protocol described here in section Materials and Methods to now consistently obtain full genomes from ≥log4 copies/ml samples, both library approaches on their own still yield a minority of viral sequences (1–5%; red inserts in **Figure 1**) relative to host and reagent background (black inserts in **Figure 1**; Luk et al., 2015; Berg et al., 2016). Using the probes described above, we explored whether target capture of HIV reads from these conventional libraries could boost NGS sensitivity for viral sequences present in low abundance (**Figure 1**).

Since xGen probes are modified with a 5′-biotin tag, Nextera XT adaptors lacking biotin needed to be synthesized to avoid streptavidin-mediated capture of all input sequences. Individually barcoded libraries were pooled together for a single capture, hybridized to HIV xGen probes on magnetic streptavidin beads, and washed to eliminate background (non-HIV) sequences. After PCR amplification, we note that the size range of resulting xGen libraries was often noticeably larger than their unenriched precursors (350–500 vs. 200–300 nt; data not

TABLE 2 | NGS data for samples sequenced only by HIV xGen.

Specimen	Country	HIV group or M subtype	Viral load (Log_{10})	Library input (copies)	Library type	Total reads	HIV reads	% HIV reads	Genome coverage%	Avg coverage depth	Standard deviation
459-16	Cameroon	HIV-1 CRF02	5.31	41,345	SSIII-xGen	8,303,330	8,043,055	96.87	100	108,492.77	214,888.09
A1699	Cameroon	HIV-1 F2	5.04	21,587	HIV SMART-xGen	5,380,619	5,336,115	99.17	100	70,586.38	117,687.56
A1786	Cameroon	HIV-1 G	4.73	10,573	HIV SMART-xGen	108,242	93,856	86.71	83	1,247.45	2,496.38
8128965‡	South Africa	HIV-1 C	4.50	6,226	HIV SMART-xGen	243,571	232,511	95.46	95	3,017.97	8,065.69
144-26†	Cameroon	HIV-1 G	4.03	2,110	HIV SMART-xGen	247,845	219,844	88.70	93	2,941.81	4,593.91
A1185	Cameroon	HIV-1 CRF01	3.87	1,459	HIV SMART-xGen	96,329	69,459	72.11	79	940.92	2,784.90
5056135	Senegal	HIV-1 C	3.67	921	HIV SMART-xGen	517,740	454,502	87.79	100	6,237.01	9,850.11
577-27‡	Cameroon	HIV-1 A	3.67	921	HIV SMART-xGen	111,404	94,003	84.38	79	1,278.21	3,069.71
42-877	Uganda	HIV-1 D	3.66	900	HIV SMART-xGen	313,662	279,983	89.26	100	3,829.33	5,495.33
112-11‡	Cameroon	HIV-1 F2	3.63	840	HIV SMART-xGen	204,740	158,815	77.57	91	2,236.33	3,634.36
P4142	DRC	HIV-1 A1	3.29	383	HIV SMART-xGen	814501	24	0.29	34	0.34	0.70
2528	DRC	HIV-1 A1	3.26	358	HIV SMART-xGen	553,492	0	0	0	0	0
PBS70-233	DRC	HIV-1 A1	3.24	342	HIV SMART-xGen	1,322	0	0	0	0	0
PBS369-87	DRC	HIV-1 A1	3.18	297	HIV SMART-xGen	947,145	30	0.31	14	0.42	1.13
475-17	Cameroon	HIV-1 A	3.00	262	HIV SMART-xGen	238,092	1,753	0.74	20	21.24	73.91
669-17	Cameroon	HIV-1 CRF22	2.93	172	SSIII-xGen	119,116	38,654	32.45	51	509.50	1584.65
129-26‡	Cameroon	HIV-1 D	2.82	173	HIV SMART-xGen	441,846	40,680	9.21	53	579.49	2,232.28
844-55	Cameroon	HIV-1 B/02	2.77	155	HIV SMART-xGen	224,916	2,542	1.13	18	28.16	121.64
5020-19	Cameroon	HIV-1 C	2.76	151	HIV SMART-xGen	602,884	1,444	0.24	12	6.80	60.30
4188-11	Cameroon	HIV-1 CRF11	2.75	148	HIV SMART-xGen	702,804	516	0.07	12	5.58	22.09
10047105267	Thailand	HIV-1 CRF01	2.72	138	HIV SMART-xGen	893,628	149,379	16.72	42	2,126.61	10,928.35
193-6	Cameroon	HIV-1 A	2.65	117	HIV SMART-xGen	370,450	496	0.13	14	3.37	14.69
961-09‡	Cameroon	HIV-1 CRF37	2.43	71	HIV SMART-xGen	229,794	197	0.09	10	1.45	9.31
363-24†	Cameroon	HIV-1 CRF13	nd	n/a	HIV SMART-xGen	133,832	117,148	87.53	85	1,478.12	4,140.94
1173-31	Cameroon	HIV-1 CRF37	nd	n/a	HIV SMART-xGen	51,448	38,288	74.42	77	524.27	1,835.32
9505343‡	Senegal	HIV-1 CRF06	nd	n/a	HIV SMART-xGen	99,834	37,446	37.51	91	484.82	945.49
814-43†	Cameroon	HIV-1 URF	nd	n/a	HIV SMART-xGen	485,500	438,989	90.42	93	5,872.10	11,973.02
770-8	Cameroon	HIV-1 URF	nd	n/a	HIV SMART-xGen	5,649,660	5,605,300	99.21	100	74,539.08	135,007.96
AB260†	Cameroon	HIV-1 group O	nd	n/a	HIV SMART-xGen	8,587,083	7,390,731	86.07	99	80,634.90	117,442.21
296	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	17,777,038	17,243,456	97.00	100	220,976.08	354,818.32
108-08	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	115,400	8,312	7.20	22	119.92	422.06
1095-04	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	255,480	77,095	30.18	52	972.67	4,214.80
1225-51	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	1,067,940	805,255	75.40	100	9,803.67	15,258.08
126-12	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	284,826	46,947	16.48	56	463.94	1,313.88
20-02	Cameroon	HIV-1 group O	nd	n/a	SSIII-xGen	308,694	174,250	56.45	94	2,183.30	3,452.81
136-16	Cameroon	HIV-2 group B	nd	n/a	SSIII-xGen	6,068	2,922	48.15	75	36.03	47.93

† *Genome completed (100%) with Sanger.*
‡ *Genome coverage increased with Sanger but still incomplete.*

shown). MiSeq runs were performed on an HIV-xGen super-library that typically multiplexed 6–26 samples, all with unique dual barcodes to permit parsing of data (**Figure 1**).

Diverse Sequences Are Captured and Enriched in HIV-xGen Libraries

To establish that the expected range of diversity was indeed captured by this method and determine whether the yield of viral reads increased, HIV xGen was applied to a variety of HIV-1 and HIV-2 strains. We began with high-titer specimens or virus isolates we previously sequenced to assess the fidelity of probes and exclude the possibility that gaps in xGen-generated coverage were a result of reads missing from the Nextera starting material. Also, by remaking libraries and arriving at the same consensus, we could demonstrate that HIV-xGen was not introducing artifacts or sequence bias. For HIV-1 group M, a

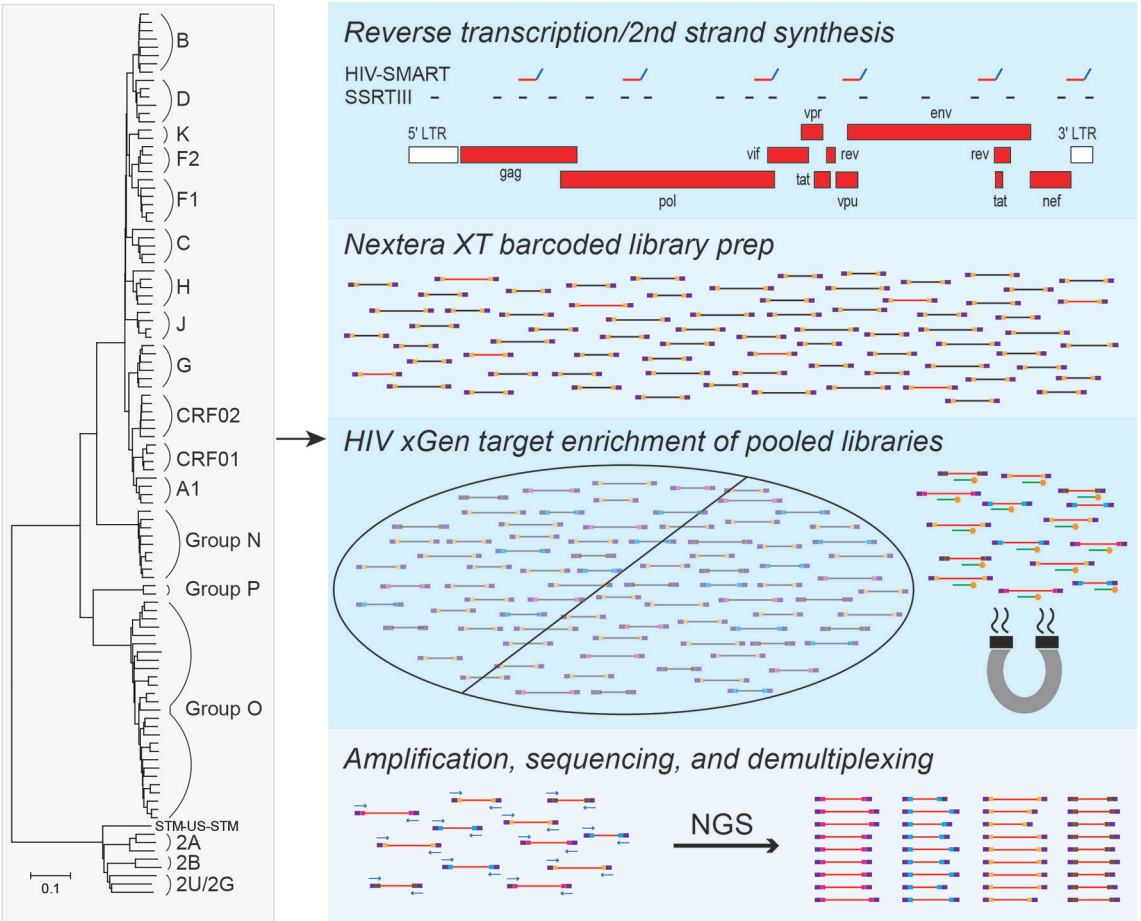


FIGURE 1 | HIV-xGen strategy. 651 probes were selected to tile all HIV-1 and HIV-2 strains present in the phylogenetic tree at 1X coverage. Reverse transcription and second strand synthesis were performed by random priming with Superscript/Sequenase or by the HIV-SMART method. Nextera XT was used to convert cDNA to barcoded Illumina libraries consisting of both HIV (red inserts) and background (black) reads. Pooled libraries were hybridized to xGen probes (green) with 5′-biotin tags (gold) for a single capture and selected by magnetic bead separation. Multiplexed libraries were amplified by universal KAPA primers, sequenced on a MiSeq, and reads were parsed by barcode.

high titer subtype C strain from South Africa (8119636; log 5.62 copies/ml) previously sequenced by HIV-SMART was remade this time by random priming (SSRTIII) and once again yielded 98% coverage (Berg et al., 2016). Following a post-Nextera HIV-xGen selection, 100% coverage was obtained, implying that sufficiently complementary sequences were present among the probes (**Figure 2**; **Table 1**). Notably, whereas only 2.69% of metagenomics reads mapped to HIV, this improved dramatically to 98.23% with HIV-xGen selection. The resulting SSIII-derived HIV-xGen consensus sequence was 99.99% identical to the HIV-SMART sequence. We continued to evaluate the probe set for the ability to capture HIV-1 groups O and N and HIV-2A sequences from virus isolate-generated libraries. Complete coverage was previously obtained in each case, with HIV-SMART libraries comprised of 3.8% (LA34; group O) and 10.0% (LA28; group N) HIV reads and randomly primed Ovation Single Cell libraries with 1.33% (LA38; HIV-2A) HIV reads (Berg et al., 2016; Yamaguchi et al., 2017). Here, cDNA libraries of each isolate were remade by random priming in this study and once again, 100% coverage was achieved for all three strains with a post-Nextera HIV-xGen enrichment, confirming that probes adequately covering these diverse groups were present and functional. The percentage of HIV reads once again increased to 97.23% for LA34 (group O), 99.39% for LA28 (group N) and 96.89% for LA38 (HIV-2A; **Figure 2A**). Likewise, the consensus sequences derived from HIV-xGen were 100%, 99.99%, and 100% identical to prior sequences.

The HIV-xGen method was then applied to additional clinical samples from Cameroon and Senegal with either viral loads ≥log5 copies/ml or those with unknown titers, this time with HIV-SMART libraries as the starting cDNA. For HIV-1 group M, a subtype F2, a CRF06, and a unique recombinant (URF) were sequenced and 100%, 91%, and 100% coverage was obtained for each, respectively (**Figure 2B**; **Table 2**). An HIV-1 group O strain (O-AB260) also yielded 99% coverage. The percentages of HIV reads in the total ranged from 86 to 99%. Thus, a diverse set of

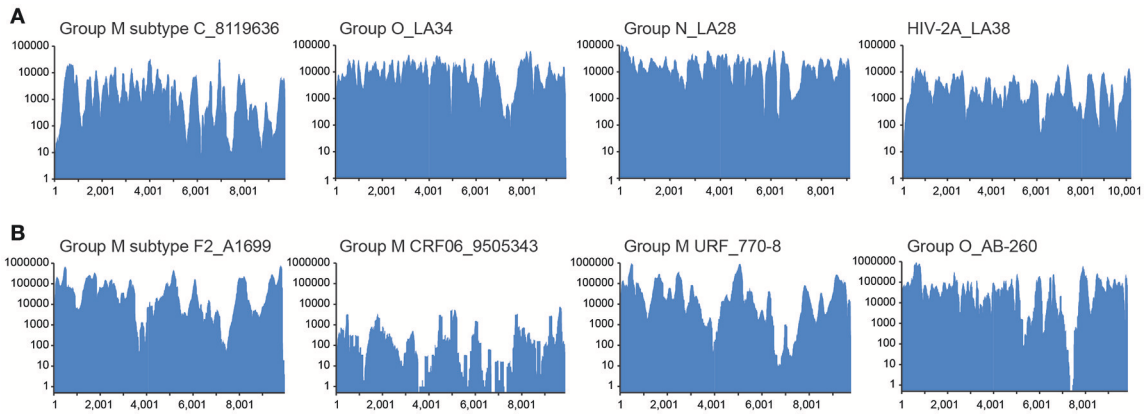


FIGURE 2 | Diverse sequences are captured and enriched in HIV-xGen libraries. **(A)** Coverage plots of SSRTIII libraries enriched by HIV-xGen that were sequenced previously: HIV-1 group M subtype C (8119636), group O (LA34), group N (LA28), and HIV-2A (LA38). **(B)** Coverage plots of new HIV-1 strains libraries generated by HIV-SMART and followed by enrichment with HIV-xGen: HIV-1 group M subtype F2 (A1699), CRF06 (9505343), URF (770-8), and group O (AB260).

high titer specimens were fully sequenced by HIV-xGen selection regardless of which cDNA synthesis method was deployed.

HIV-xGen Dramatically Increases Sensitivity for Low Titer Specimens

The value of HIV-xGen will reside in its ability to fully sequence low titer specimens while multiplexing to the same or greater extent. Representative results from strains with viral loads of log 4.59 copies/ml (PBS1342; URF), log 3.86 copies/ml (PBS1191; subtype G) and log 3.47 copies/ml (PBS888; subtype A) demonstrate the dramatic improvements in coverage with enrichment compared to without (**Figure 3A**). For HIV SMART, genome coverage increases from 86, 51, and 56% without xGen to 95%, 100%, and 100% with xGen, respectively. Similarly, genome coverage with Superscript (SSRTIII) changes from 99, 45, and 54% without xGen to 97, 96, and 96% with xGen, respectively. Here again, HIV-xGen libraries were comprised almost entirely (90–99%) of HIV sequence regardless of the cDNA synthesis method chosen (SMART or SSRTIII) and resulting consensus sequences were 99.51, 99.04, and 99.76% identical, respectively.

Thirteen additional samples from the Democratic Republic of Congo, ranging in titers of log 3.58 to 5.78 copies/ml, were sequenced here by the HIV-SMART ± xGen method (**Table 1**). These specimens together with the examples above reveal a median 1,147x (range 24.7–56, 509x) boost in HIV read yield upon xGen enrichment. For high titer specimens, percent genome coverage is largely unaffected (see below **Figure 3C**), although the depth of coverage is substantially increased as the same reads are re-sequenced. However, specimens <log 4.5 copies/ml saw both a significant increase in depth and overall genome coverage, indicating that additional HIV reads are present in libraries which have not been sequenced without enrichment (**Table 1**, **Figure 3C**). In a few instances, fewer than 10 reads were initially mapped, which following xGen selection, resulted in 97–100% coverage (subtype K, P3844; subtype A1, PBS0724). The average increase in percent genome coverage for xGen-enriched compared to unenriched samples was 40.5%. Indeed, all samples >log 3.5 copies/ml had ≥79% coverage, with the majority (72%) of these having >95% genome coverage. Once again, strain consensus sequences were virtually identical independent of xGen, as well as when PCR duplicate reads were removed during mapping (**Supplemental Table S2A**). Similarly, the total number of minor variants (MV; 10–50%) detected was consistent between methods. However, while PCR duplicate removal had minimal effect on the extent to which the exact same MVs were detected (70–100% overlap), this overlap was reduced when comparing (-)xGen to (+)xGen datasets (28–84% overlap; **Supplemental Table S2B**).

We next explored strains with viral loads ranging from log 2.6–3.6 copies/ml, continuing with the HIV-SMART + HIV xGen approach (**Table 2**). Coverage plots for xGen-enriched libraries in **Figure 3B** illustrate that while full genomes are not possible with this method in this titer range, the partial coverage obtained for some can still be substantial (e.g., 40–80%). Once again, sequences from different geographies were successfully captured, including a subtype F2 (112-11, Cameroon), D (129-26; Cameroon), and CRF01 (10047105267; Thailand). As testament to the method sensitivity, an HIV-2 strain (136-16; Cameroon), which typically replicates at low titers, was extracted from diluted patient plasma and attained 75% genome coverage (**Figure 3B**). The average percent genome sequenced for those with ≤log 3 copies/ml is 26 ± 18%. For two samples at log 3.24 and 3.26 copies/ml, zero HIV reads were obtained. It is noteworthy that input for cDNA synthesis in this range is <200 copies of virus and that for many of these samples, simply obtaining a product by RT-PCR was also a challenge.

To summarize the results for all samples attempted in which a titer was known, genome coverage was plotted against viral load. Most samples >log 3.5 copies/ml achieved near-complete coverage and those below this threshold yielded a partial genome (**Figure 3C**). Different colors (±xGen) for the same samples (identical viral loads) illustrate the significant jumps in coverage following enrichment. There were several strains for which low sample volume precluded viral load testing, particularly for

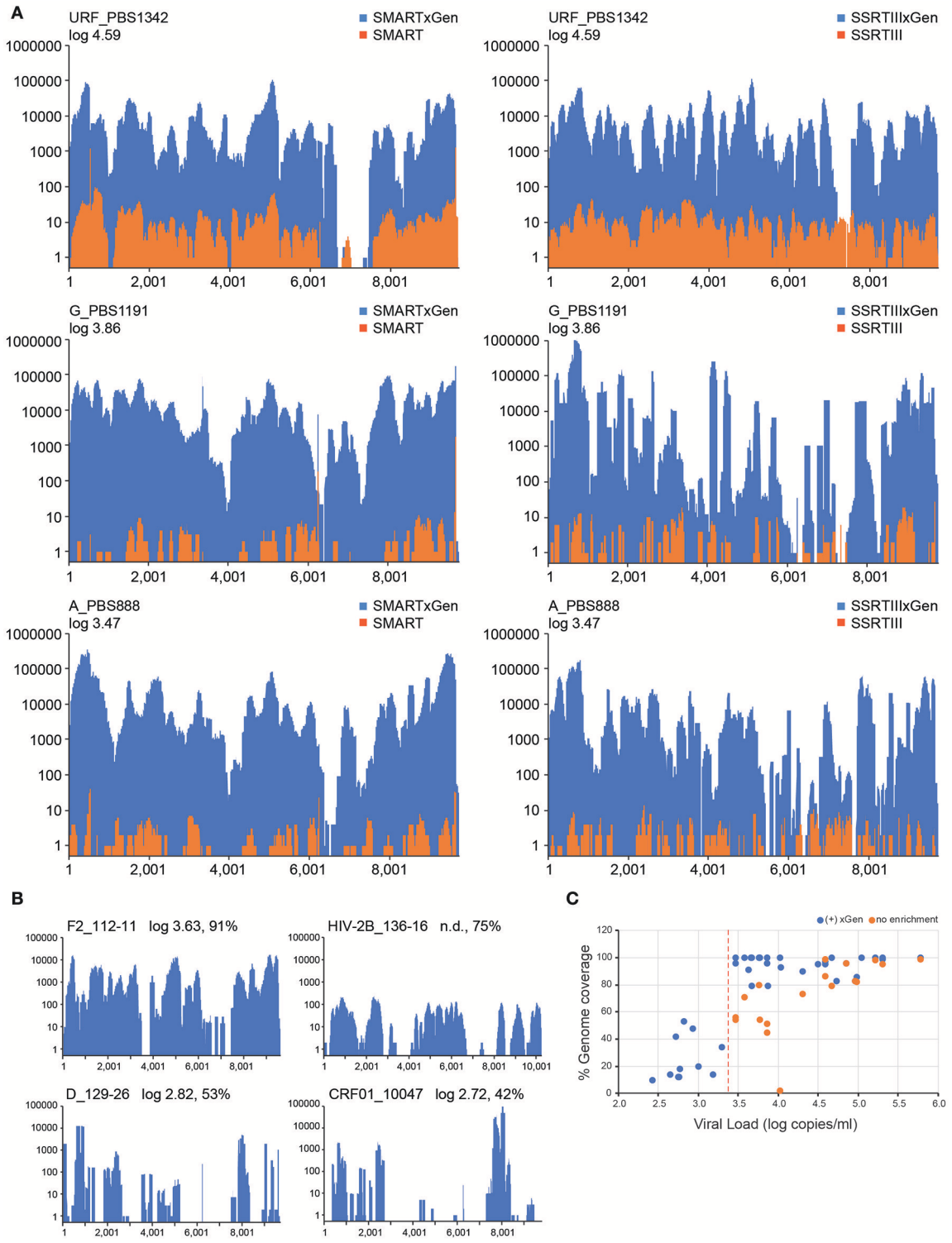


FIGURE 3 | A HIV-xGen dramatically increases sensitivity for low titer specimens. **(A)** Coverage plots are shown for HIV-SMART libraries (top panels) without (orange) and with (blue) HIV-xGen as well as Superscript libraries without (orange) and with (blue) HIV-xGen. PBS1342 is a URF, PBS1191 is subtype G and PBS888 is subtype A. **(B)** Coverage plots for strains with titers ranging from log 2.7–3.6 copies/ml including a subtype F2 (112-11), a subtype D (129-26), a CRF01 (10047105267) and an HIV-2 strain (136-16) of unknown viral load. **(C)** Plot of viral load vs. genome coverage for all new strains sequenced with a known titer. Red dashed line = log 3.4 copies/ml; orange dots = no enrichment; blue dots = xGen enrichment.

HIV-1 group O and rare circulating recombinants (**Table 2**). For more than half of these we succeeded in obtaining >90% of the genome.

HIV xGen Facilitates Classification and Characterization of Diverse HIV Strains

A total of 50 new strains originating from the Democratic Republic of Congo, Cameroon, Thailand, South Africa, Senegal, and Uganda were sequenced by HIV-xGen. Phylogenetic classification of 28 complete or near-complete (>90%) genomes determined in this study are shown using a 6,252 nt gap-stripped alignment (**Figure 4A**). Many of the major subtypes and CRFs are represented here, as well as four group O strains, illustrating the breadth of viral diversity captured by this method. Phylogenetic classifications of partial genome sequences with 75–90% (**Supplemental Figure S1**) and <75% (**Supplemental Figure S2**) coverage are found in the **Supplemental Information**.

Strains branching basal to Group M subtype nodes were investigated further by Simplot and boot-scanning to reveal evidence of recombination. For each initial Simplot analysis preceding the final bootscan shown, appropriate references were included to verify that the strain in question was more similar to the recombinant sequence than to the pure reference sequence of the same subtype (e.g., A, G, etc.). Sub-genomic RT-PCR of *env* immunodominant region (IDR) originally categorized 770-8 as a CRF13. While NGS confirmed the classification of this portion of the sequence, we were able to determine with the full genome that it was actually a unique recombinant form consisting of CRF02, CRF37 and CRF13 sequences. Similarly, full genome characterization of 814-43 demonstrated it was not simply a CRF06, but rather a URF consisting of CRF06 and CRF02 sequence (**Figure 4B**). Plotting consensus base call percentages at each position ruled out dual or super-infections, since aside from the occasional minor variant, values approached 100% throughout the genome. No continuous stretches of lower consensus base call percentages were observed, indicating only one major recombinant species was present (data not shown).

DISCUSSION

HIV-xGen is a universal, robust, and cost-effective back-end to any cDNA method deployed for next generation sequencing of HIV-1 and HIV-2. Previously, with benzonase-treated extractions and either our optimized gene-specific (HIV-SMART) or standard random priming approaches for cDNA synthesis, we were still challenged by sensitivity. While we obtained much greater coverage for samples with titers between log 4 and log 5 copies/ml, genomes were still incomplete (Berg et al., 2016; Rodgers et al., 2017a). Now, with HIV-xGen, we can routinely obtain full genomes at a lower limit of log 3.5 copies/ml, whereas previously without enrichment, samples in the log 3.5–4.5 range would only yield 20–50% coverage. Thus, it was clear that these HIV reads were actually present in libraries, but we were not sequencing to a sufficient read depth to observe them. Below this log 3.5 threshold, the ability to adequately sequence samples is likely a limitation of cDNA synthesis, and not xGen; it cannot capture and amplify material that was never reverse transcribed. With specimens each having inherent differences in host background, detection is not linear in this range. Some may obtain >50% coverage whereas others of similar titer recover no sequence at all. Overall, our results are consistent with what others have reported for probe-mediated positive selection of viral sequences (Bonsall et al., 2015; Briese et al., 2015). Simply sequencing deeper, on a higher throughput instrument (e.g., HiSeq vs. MiSeq), or multiplexing less could certainly provide improved detection and coverage of low titer samples, but the overall read proportions would likely remain the same. Now with xGen, far more samples can be processed and sequenced at once as a greater percentage of reads are viral, saving time and resources. Throughput and NGS capacity are further increased by virtually eliminating the sequencing of host background.

As with HIV-SMART, the objective with HIV-xGen was to fully characterize strains for the purpose of surveillance. With the sensitivity we now demonstrate, the number of specimens previously deemed too challenging to sequence by NGS due to low titer has markedly declined. As a greater proportion of patients are on therapy and able to suppress viral loads, it is imperative that our methods can adequately characterize low titer specimens. While we only attempted to multiplex a maximum of 26 samples at a time, the high percentage of HIV reads from the total per barcode suggests many more libraries (e.g., an entire 96-well plate) could be pooled without detriment. Considering the depth of coverage now possible, assessing levels of minor variants and quasi-species in samples could be an attractive application for HIV-xGen. For example, detection of minor variants at clinically relevant levels (e.g., >10%) to predict the emergence of drug resistance and treatment failure should be readily achievable with this method (Li and Kuritzkes, 2013; Obermeier et al., 2014; Noguera-Julian et al., 2017). However, primer IDs controlling for starting cDNA populations and potential PCR bias were not used in this study and the original proportions of minor variants after Nextera amplification might be expected to drift further after additional rounds of HIV-xGen amplification (Jabara et al., 2011; Boltz et al., 2016). Indeed, we showed that while consensus sequences do not change with enrichment or inclusion of duplicate reads, minor variant populations are affected by xGen (**Supplemental Table S2**). Nevertheless, our primary goal was to not limit interrogation to sub-genomic regions via amplicon sequencing, but rather to better comprehend the complete extent of diversity in entire genomes found in different geographic regions at different times. As an example, our prior surveillance efforts in Cameroon have consistently observed a predominance of CRF02, CRF06, CRF13, and CRF37 (Rodgers et al., 2017b). Now it appears that the URFs we are detecting in this region are recombinants and contain genome segments that do not phylogenetically cluster with homologous sequences derived from any of the classified HIV-1M (**Figure 4**).

With the batching of samples during hybridization and amplification steps, the inability to control for the resulting proportion of reads/barcode is one shortcoming of the method. Generally, the resulting read numbers were a reflection of the starting viral loads. Thus, we recommend, if possible, that samples be grouped by titer to avoid an imbalance. Nevertheless,

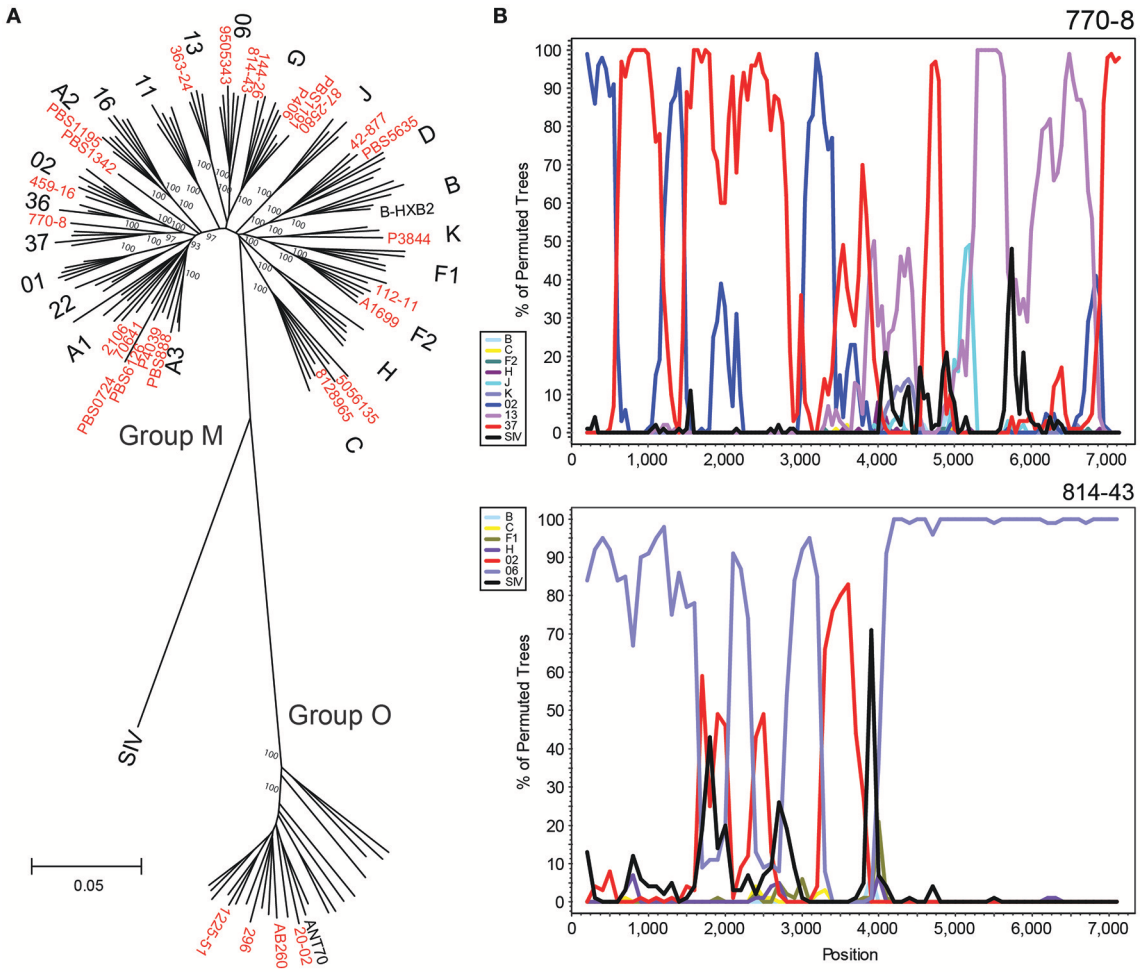


FIGURE 4 | HIV-xGen enables full genome characterization of rare and diverse strains. **(A)** Neighbor joining phylogenetic tree of 28 new full genomes (red) generated from a 6,252 nt gap-stripped alignment. Specimens sequenced in this study are shown in red. **(B)** Boot-scanning analysis of unique recombinant forms 770-80 and 814-43.

given the potential for one or more samples to predominate in a run relative to the others in the pool, cross-contamination of reads between barcodes on the MiSeq was of concern (Lee et al., 2016). Therefore, raw data from each sample was mapped to each individual HIV xGen-derived consensus sequence of other samples in the same run to identify regions of perfect identity which might indicate the incorrect binning of reads. Fortunately, this was infrequently observed, with most errant reads derived from high read depth samples, and particularly in those samples where a barcode may have been shared. Thus, we further recommend that unique dual barcode pairs be chosen for each library in a pool. Here, reads clearly originating from another sample were removed from the dataset and the mapping was repeated to derive an accurate consensus. In most cases, we had prior Sanger data to compare the NGS consensus sequences against to verify that the final sequence obtained was correct.

Another shortcoming of the method is that the full scope of HIV diversity cannot be known, therefore probes comprehensive of all sequences cannot be designed, and yet this is an argument in favor of continued surveillance. As an example in **Figure 3A**, PBS1342 exhibits a gap in the *env* region using either cDNA synthesis approach. The HIV-SMART reverse primer binding site is considerably downstream of the gap and fails to explain the absence of sequence, irrespective of enrichment. In contrast, randomly primed libraries did sequence successfully across this region, for which only a small portion (nt 7,196–7,455) was not recovered by xGen. Comparison of the 120 nt probes spanning this region to the PBS1342 consensus sequence revealed an overall lower identity (78–85%), but presumably the concentration of mismatches and indels we observed were the major factor. In the first probe, 14 of the total 18 mismatches were focused in the 3′ 50 nucleotides, the middle probe had two insertions (6, 3 nt), one deletion (3 nt), and 13 mismatches all situated in the 5′ 53 nucleotides, and in the third there were 19 mismatches over a stretch of 33 nucleotides, preceded by two insertions (6, 27 nt) and a deletion (15 nt). Adequate coverage of envelope will likely remain a challenge, particularly as we explore geographies with high sequence diversity such as the DRC, yet

the tolerance for mismatches displayed throughout the genomes of numerous strains from HIV-1 and HIV-2 groups speaks to the robustness of the method.

It is essential that diagnostic tests keep pace with HIV and other rapidly mutating viruses by proactively seeking out strains in circulation that may escape detection with current assays (Brennan et al., 2006). The ability to fully characterize multiple samples simultaneously without regard for subtype, group, or titer opens up greater opportunities for future surveillance. At the same time, using HIV-xGen to retrospectively examine archived specimens and understand the origins of the HIV epidemics is an equally compelling application (Rodgers et al., 2017a). Metagenomic NGS combined with selective viral sequencing promise a new era in diagnostic virology (Barzon et al., 2011; Quiñones-Mateu et al., 2014; Kumar et al., 2017). As sensitivity and throughput increases with methods like HIV-xGen, we are one step closer to realizing this potential.

AUTHOR CONTRIBUTIONS

JY and MB conceived of methodology and designed HIV-xGen probes. JY, AO, and KF performed experiments and data analysis. MR performed data analysis, assisted with experimental design, and reviewed manuscript. MB performed data analysis, directed experimental design, wrote manuscript and made figures. GC assisted experimental design and reviewed manuscript. TQ and OL collected and characterized DRC samples and reviewed manuscript. NN, DM, and LK collected and characterized Cameroon samples and reviewed manuscript.

FUNDING

This study was funded by Abbott Laboratories.

ACKNOWLEDGMENTS

We thank the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases, NIH, for support in obtaining and providing the samples from Project SIDA and scientific input. We thank Cameroon personnel Bih Awazi, Jules Bertrand, and Kenmegne Sidje for their assistance, and we acknowledge the National Blood Centre of the Thai Red Cross Society in Bangkok, Thailand for providing specimens for this study. We thank Nicholas Downey at IDT for assistance with xGen probe design and Nextera adaptor synthesis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02150/full#supplementary-material

REFERENCES

Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., et al. (2012). Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. *PLoS ONE* 7:e49602. doi: 10.1371/journal.pone.0049602

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., and Palù, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* 12, 7861–7884. doi: 10.3390/ijms12117861

Berg, M. G., Lee, D., Coller, K., Frankel, M., Aronsohn, A., Cheng, K., Forberg, K., et al. (2015). Discovery of a novel human pegivirus in blood associated with hepatitis C virus co-infection. *PLoS Pathog.* 11:e1005325. doi: 10.1371/journal.ppat.1005325

Berg, M. G., Yamaguchi, J., Alessandri-Gradt, E., Tell, R. W., Plantier, J. C., Brennan, CA. (2016). A pan-HIV strategy for complete genome sequencing. *J. Clin. Microbiol.* 54, 868–882. doi: 10.1128/JCM.02479-15

Bimber, B. N., Dudley, D. M., Lauck, M., Becker, E. A., Chin, E. N., Lank, S. M., et al. (2010). Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. *J. Virol.* 84:12087–12092. doi: 10.1128/JVI.01378-10

Boltz, V. F., Rausch, J., Shao, W., Hattori, J., Luke, B., Maldarelli, F., et al. (2016). Ultrasensitive single-genome sequencing: accurate, targeted, next generation sequencing of HIV-1 RNA. *Retrovirology* 13:87. doi: 10.1186/s12977-016-0321-6

Bonsall, D., Ansari, M. A., Ip, C., Trebes, A., Brown, A., Klenerman, P., et al. (2015). ve-SEQ: robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Res* 4:1062. doi: 10.12688/f1000research.7111.1

Brennan, C. A., Bodelle, P., Coffey, R., Devare, S. G., Golden, A., Hackett J. Jr., et al. (2008). The prevalence of diverse HIV-1 strains was stable in cameroonian blood donors from 1996 to 2004. *J. Acquir. Immune Defic. Syndr.* 49, 432–439. doi: 10.1097/QAI.0b013e31818a6561

Brennan, C. A., Bodelle, P., Coffey, R., Harris, B., Holzmayer, V., Luk, K. C., et al. (2006). HIV global surveillance: foundation for retroviral discovery and assay development. *J. Med. Virol.* 78, S24–S29. doi: 10.1002/jmv.20603

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., et al. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* 6, e01491–e01415. doi: 10.1128/mBio.01491-15

Cohen, J. (1997). The rise and fall of Project SIDA. *Science* 278, 1565–1568. doi: 10.1126/science.278.5343.1565

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., et al. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* 5:16532. doi: 10.1038/srep16532

Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., et al. (2012). Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* 50, 3838–3844. doi: 10.1128/JCM.01516-12

Giallonardo, F. D., Töpfer, A., Rey, M., Prabhakaran, S., Duport, Y., Leemann, C., et al. (2014). Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res.* 42:e115. doi: 10.1093/nar/gku537

Hall, R. J., Wang, J., Todd, A. K., Bissielo, A. B., Yen, S., Strydom, H., et al. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods* 195, 194–204. doi: 10.1016/j.jviromet.2013.08.035

Hemelaar, J. (2013). Implications of HIV diversity for the HIV-1 pandemic. *J. Infect.* 66, 391–400. doi: 10.1016/j.jinf.2012.10.026

Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A., and Swanstrom, R. (2011). Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20166–20171. doi: 10.1073/pnas.1110064108

Kumar, A., Murthy, S., and Kapoor, A. (2017). Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Res.* 239, 172–179. doi: 10.1016/j.virusres.2017.06.005

Lee, H. K., Lee, C. K., Tang, J. W., Loh, T. P., and Koay, E. S. (2016). Contamination-controlled high-throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer. *Sci. Rep.* 6:33318. doi: 10.1038/srep33318

Li, J. Z., and Kuritzkes, D. R. (2013). Clinical implications of HIV-1 minority variants. *Clin. Infect. Dis.* 56,1667–1674. doi: 10.1093/cid/cit125

Luk, K. C., Berg, M. G., Naccache, S. N., Kabre, B., Federman, S., Mbanya, D., et al. (2015). Utility of metagenomic next-generation sequencing for characterization of hiv and human pegivirus diversity. *PLoS ONE* 10:e0141723. doi: 10.1371/journal.pone.0141723

Miyazato, P., Katsuya, H., Fukuda, A., Uchiyama, Y., Matsuo, M., Tokunaga, M., et al. (2016). Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome. *Sci. Rep*. 6:28324. doi: 10.1038/srep28324

Noguera-Julian, M., Edgil, D., Harrigan, P. R., Sandstrom, P., Godfrey, C., Paredes, R., et al. (2017). Next-generation human immunodeficiency virus sequencing for patient management and drug resistance surveillance. *J. Infect. Dis*. 216, S829–S833. doi: 10.1093/infdis/jix397

Obermeier, M., Ehret, R., Wienbreyer, A., Walter, H., Berg, T., Baumgarten, A. (2014). Resistance remains a problem in treatment failure. *J. Int. AIDS. Soc*. 17:19756. doi: 10.7448/IAS.17.4.19756

Oster, A. M., Switzer, W. M., Hernandez, A. L., Saduvala, N., Wertheim, J. O., Nwangwu-Ike, N., et al. (2017). Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013. *Ann. Epidemiol.* 27, 244–251 e241. doi: 10.1016/j.annepidem.2017.02.002

Pyne, M. T., Hackett, J. Jr., Holzmayer, V., and 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

Quiñones-Mateu, M. E., Avila, S., Reyes-Teran, G., and Martinez, M. A. (2014). Deep sequencing: becoming a critical tool in clinical virology. *J. Clin. Virol*. 61, 9–19. doi: 10.1016/j.jcv.2014.06.013

Redd, A. D., Mullis, C. E., Serwadda, D., Kong, X., Martens, C., Ricklefs, S. M., et al. (2012). The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J. Infect. Dis*. 206, 267–274. doi: 10.1093/infdis/jis325

Rodgers, M. A., Vallari, A. S., Harris, B., Yamaguchi, J., Holzmayer, V., Forberg, K., et al. (2017b). Identification of rare HIV-1 Group, N., HBV AE, and HTLV-3 strains in rural South Cameroon. *Virology* 504, 141–151. doi: 10.1016/j.virol.2017.01.008

Rodgers, M. A., Wilkinson, E., Vallari, A., McArthur, C., Sthreshley, L., Brennan, C. A., et al. (2017a). Sensitive next-generation sequencing method reveals deep genetic diversity of HIV-1 in the democratic republic of the congo. *J. Virol*. 91:e01841-16. doi: 10.1128/JVI.01841-16

Semaille, C., Barin, F., Cazein, F., Pillonel, J., Lot, F., Brand, D., et al. (2007). Monitoring the dynamics of the HIV epidemic using assays for recent infection and serotyping among new HIV diagnoses: experience after 2 years in France. *J. Infect. Dis*. 196, 377–383. doi: 10.1086/519387

Swanson, P., Harris, B. J., Holzmayer, V., Devare, S. G., Schochetman, G., Hackett J. Jr., et al. (2000). Quantification of HIV-1 group M (subtypes A-G) and group O by the LCx HIV RNA quantitative assay. *J. Virol. Methods* 89, 97–108. doi: 10.1016/S0166-0934(00)00205-6

Swenson, L. C., Däumer, M., and Paredes, R. (2012). Next-generation sequencing to assess HIV tropism. *Curr. Opin. HIV AIDS* 7, 478–485. doi: 10.1097/COH.0b013e328356e9da

Yamaguchi, J., Brennan, C. A., Alessandri-Gradt, E., Plantier, J. C., Cloherty, G. A., Berg, M. G. (2017). HIV-2 surveillance with next-generation sequencing reveals mutations in a cytotoxic lymphocyte-restricted epitope involved in long-term nonprogression. *AIDS Res. Hum. Retroviruses* 33, 347–352. doi: 10.1089/aid.2016.0229

Conflict of Interest Statement: JY, MR, AO, KF, GC, and MB are all Abbott employees and shareholders.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 October 2018
doi: 10.3389/fmicb.2018.02358

Characterization of Inducible Transcription and Translation-Competent HIV-1 Using the RNAscope ISH Technology at a Single-Cell Resolution

*Wang Zhang[1,2], Sara Svensson Akusjärvi[1], Anders Sönnerborg[1,3] and Ujjwal Neogi[1]**

[1] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Solna, Sweden, [2] Science for Life Laboratory, Division of Proteomics and Nanobiotechnology, KTH Royal Institute of Technology, Solna, Sweden, [3] Department of Medicine Huddinge, Unit of Infectious Diseases, Karolinska Institutet, Karolinska University Hospital, Solna, Sweden

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Kumitaa Theva Das,
Advanced Medical and Dental Institute, Malaysia
Dorothy Ellen Lewis,
McGovern Medical School, United States

***Correspondence:**
Ujjwal Neogi
ujjwal.neogi@ki.se

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *16 July 2018*
Accepted: *14 September 2018*
Published: *02 October 2018*

Citation:
Zhang W, Svensson Akusjärvi S, Sönnerborg A and Neogi U (2018) Characterization of Inducible Transcription and Translation-Competent HIV-1 Using the RNAscope ISH Technology at a Single-Cell Resolution. Front. Microbiol. 9:2358. doi: 10.3389/fmicb.2018.02358

Identifying the source and dynamics of persistent HIV-1 at single-cell resolution during cART is crucial for the design of strategies to eliminate the latent HIV-1 reservoir. An assay to measure latent HIV-1 that can distinguish inducible from defective proviruses with high precision is essential to evaluate the efficacy of HIV-1 cure efforts but is presently lacking. The primary aim of this study was therefore to identify transcription and translation competent latently infected cells through detection of biomolecules that are dependent on transcriptional activation of the provirus. We investigated the applicability of two commercially available assays; PrimeFlow™ RNA Assay (RNAflow) and RNAscope® ISH (RNAscope) for evaluation of the efficacy of latency reversal agents (LRAs) to reactivate the HIV-1 latent reservoir. The J-Lat cell model (clones 6.3, 9.3, and 10.6) and four LRAs was used to evaluate the sensitivity, specificity, and lower detection limit of the RNAflow and RNAscope assays for the detection and description of the translation-competent HIV-1 reservoir. We also checked for HIV-1 subtype specificity of the RNAscope assay using patient-derived subtype A1, B, C, and CRF01_AE recombinant plasmids following transfection in 293T cells and the applicability of the method in patient-derived peripheral blood mononuclear cells (PBMCs). The lower detection limit of RNAflow was 575 HIV-1 infected cells/million and 45 cells/million for RNAscope. The RNAscope probes, designed for HIV-1B, also detected other subtypes (A1, B, C, and CRF01_AE). RNAscope was applicable for the detection of HIV-1 in patient-derived PBMCs following LRA activation. In conclusion, our study showed that RNAscope can be used to quantify the number of directly observed individual cells expressing HIV-1 mRNA following LRA activation. Therefore, it can be a useful tool for characterization of translation-competent HIV-1 in latently infected cell at single-cell resolution in the fields of HIV-1 pathogenesis and viral persistence.

Keywords: RNAscope, RNAflow, HIV-1 latency, latency-reversing agents, single-cell characterization

INTRODUCTION

From a clinical perspective, three read-outs in HIV-cure research can be considered: latency reversal, reservoir reduction and viral remission after cessation of combinational antiretroviral therapy (cART) (Churchill et al., 2016). Understanding the size of the reservoir and how best to target it, is a stepping-stone for HIV-cure research. Identifying the source and dynamics of persistent HIV-1 at single-cell resolution during cART is crucial for understanding the barriers for a functional HIV cure. A primary challenge in HIV-cure research is development of robust assays that can quantify the true reservoir of integrated, replication-competent provirus within the host cellular DNA and strategies that can eliminate or control this persistent viral reservoir.

An assay to measure HIV-1 latency that can distinguish inducible from defective proviruses with high precision is essential for evaluation of efficacy of eradication efforts but is presently lacking. Quantitative HIV-1 DNA PCR assays measure all proviruses, of which the vast majority are defective (Ho et al., 2013), thereby overestimating the size of the latent replication-competent reservoir. The quantitative viral outgrowth assay (Q-VOA), termed as the "golden-standard", measures induced proviruses (Karlsson et al., 2001; Laird et al., 2013); however, it underestimates the size of the reservoir as it is unable to detect intact non-induced provirus (Eriksson et al., 2013). The technique TILDA (Tat/rev Induced Limiting Dilution Assay) measures the frequency of cells with inducible, multiply-spliced HIV-1 RNA that is absent in latently infected cells but can be induced upon viral reactivation (Procopio et al., 2015). However, large discrepancies are found between these assays that results in a significant hurdle for clinical trials that aim to measure the efficacy of HIV-1 eradication strategies (Bruner et al., 2015).

Recent studies have shown that the RNA-flow fluorescent *in situ* hybridization (FISH) technique can be useful to measure the translation or transcription-competent reservoir with high sensitivity and specificity (Baxter et al., 2017; Grau-Exposito et al., 2017). The method can also provide relevant information when studying HIV pathogenesis, persistence, and reactivation. Further, a more recent study using the branched DNA *in situ* hybridization technology combined with multiplex immunofluorescent cell-based detection of DNA, RNA, and Protein (MICDDRP) targeting both HIV-1 RNA, integrated provirus and viral proteins indicated that the method can be applied at single-cell resolution (Puray-Chavez et al., 2017).

In this study, we investigated the applicability of two commercially available assays, PrimeFlow™ RNA Assay (herein RNAflow) and RNAscope® ISH technique (herein RNAscope) for evaluation of the efficacy of latency reversal agents (LRAs) to induce the HIV-1 latent reservoir. For this purpose, the J-Lat cell model and low input of patients' peripheral blood mononuclear cells (PBMCs) were used, respectively. We also checked the HIV-1 subtype specificity of the assays. Our study indicated that due to the loss of a large number of cells in RNAflow, the RNAscope technique perform better than RNAflow. In addition, RNAscope had a lower detection limit while scanning 1×10^6 cells, independent on the patient-derived HIV-1 subtype (HIV-1A1, HIV-1B, HIV-1C, 01_AE).

MATERIALS AND METHODS

Optimization of Latency-Reversing Agents (LRAs) to Achieve Maximal Reactivation of HIV-1 Latency

To optimize the reversal capacity of different LRAs, three J-Lat cell models of HIV latency (J-Lat Full Length Clones 6.3, 9.2, 10.6, NIH AIDS Reagent Program, United States) were used (Jordan et al., 2003). These cells are modified to contain integrated GFP replacing *nef* within the proviral genome. The parental cell line, Jurkat was used as negative control (Weiss et al., 1984). Cells were cultured in RPMI medium (10% fetal bovine serum (FBS, Gibco, Life Technologies, United States), 0.2% Penicillin-Streptomycin (Penstep, Gibco, Life Technologies, United States) and 0.2% Plasmocin™ prophylactic (InvivoGen, United States) and stimulated using individual and combinations of LRAs. The LRAs were tested in each cell model at final concentration of the Protein kinase C (PKC) agonists 12-deoxyphorbol-13-acetate (Prostratin, 6 μM, Sigma-Aldrich, MO, United States); human TNFα (hTNFα; 10 ng/mL, Thermo Fisher, United States); Calcium ionophore (Ionomycin, 1.25 μM, Sigma-Aldrich, MO, United States) and Suberoylanilide hydroxamic acid (SAHA, 6.25 μM, Sigma-Aldrich, United States), known to inhibit Histone deacetylase (HDACi). The exposure time of LRAs was standardized to 48 h. Reactivation of latent provirus and cell viability were monitored using GFP expression and viability dye (LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit, Thermo Fisher, United States) using flow cytometry (BD FACSVerse™, United States).

Surface Modified Slides and Coverslips

To prevent de-attachment of cells from coverslips during RNAscope, we used different surface modified slides, e.g., H-12-Collagen, H12-PDL, GG-12-Laminin, GG-12-Fibronectin, GG-12-Gelatin (Neuvitro Corporation, Vancouver, United States) and poly-L-lysin coverslips (Corning, United States). The poly-L-lysin coverslips provided better attachment and fluorescence signal during imaging and were therefore used for RNAscope.

HIV-1 mRNA Detection Using RNAflow

Cells were exposed for 48 hrs using LRAs followed by PrimeFlow™ (PrimeFlow RNA assay, Affymetrix/eBioscience, United States) detection of HIV-1 mRNA by flow cytometry, according to the manufacturer's protocol. Forty probe pairs were used, 20 targeting *gag* and 20 targeting *pol*. Samples were collected in tubes followed by fixation and permeabilization. Probes were diluted and hybridized to the target mRNA for 2 h at 40°C. Samples were washed to remove excess probes, and signal amplification was achieved by sequential incubations with the pre-amplification and amplification mix (1.5 h at 40°C each). Amplified mRNA was labeled with Alexa Fluor

647-tagged probes (1 h at 40°C). Samples were acquired on a FACSVerse (BD Bioscience, United States) and analysis performed using FlowJo (Treestar, V10, United States). Unactivated/healthy control samples were used as gating control. Normalized positive cell per million (NPM) were calculated by the formula: NPM = (RNAflow$^+$ cells/DAPI$^+$ cells) $\times$ 10^6. All the experiments were performed at least two technical replicates with three biological replicates.

HIV-1 mRNA Detection Using RNAscope

RNAscope® (RNAscope® Fluorescent Multiplexed reagent kit, Advanced Cell Diagnostics, United States) was used as per the manufacturer's protocol and adjusted for dual detection of mRNA and protein while retaining maximal amount of cells/coverslip. The probe set, V-HIV1-clade B-C3 (Advanced Cell Diagnostics, United States), consisted of 20 dual probes targeting different segments within the *gag-pol* region. During all preparation stages, coverslips were handled gently using forceps with no high velocities upon the coverslips as it greatly affected the attachment of cells. After activation, cells were washed two times 5 min using PBS at 1500 rpm prior attachment of 500,000 cells in 100 μL on Bio-coat pre-coated poly-L-lysin coverslips (Corning, United States) 30 min at 37°C. Unattached cells were removed using PBS prior fixation using Formalin Solution, neutral buffered 10%, (NBF, Sigma-Aldrich, United States) for 20 min at RT and washed two times 5 min PBS. Pre-treatment for storage was performed by dehydration using EtOH, 50, 70, and 100% for 2 min at room temperature subsequently for storage in 100% EtOH at −20°C for up to 1 month. Storage beyond 1 month decreases the signal. For re-hydration, the opposite concentration decrease of EtOH was performed with a final 10 min equilibrium in PBS. Cells were permeabilized using 0.1% PBS Triton-x for 10 min at RT prior washing and storing coverslips in PBS, 1 min. Coverslips were attached on Superfrost glass slides (Thermo Fisher, United States) using nail polish, and a hydrophobic barrier created using Immedge™ Hydrophobic Barrier Pen (Vector Laboratory, United States). Probe hybridization was achieved by incubation of 35 μL mRNA target probes for 2 h at 40°C using a HyBez oven. The signal was amplified by subsequent incubation of Amp-1, Amp-2, Amp-3 and Amp4 (Alexa Fluor 647-tagged probe), one drop each for 30, 15, 30, and 15 min respectively at 40°C using a HyBez oven. Each incubation step was followed by two times 2 min wash using RNAscope washing buffer in slide holders with agitation (50 rpm). Glass slides were always applied into slide holder containing washing buffer, cells upward using forceps. Nucleic acids were stained using manufacturers supplied 4′,6′-diamino-2-phenylindole (DAPI, Advanced Cell Diagnostics, United States) for 30 s and washed two times PBS, if not followed by subsequent protein staining, as described below, prior nucleic acid stain. The coverslips were mounted on Superfrost™ Plus Adhesion Slides (Thermo Fisher, MA) using Prolong Gold Antifade Mountant (Thermo Fisher, United States). Unless otherwise stated, big field images of the whole coverslip were taken with a Nikon inverted confocal microscope equipped with ×20/0.75 objective and high-speed camera (Andor Zyla 4.2+, Belfast, Northern Ireland) utilizing the tiling function. The excitation/emission bandpass wavelengths used to detect DAPI, GFP/FITC and Alexa 647 were set to 405/425–480, 488/503–555 and 647/655–705 nm, respectively. Using the latest NIS Element Software with the function for object identification and automatic counting, HIV$^{mRNA+/GFP+}$ cells were identified. Normalized positive cell per million (NPM) were calculated by the formula: NPM = (RNA$^+$ cells /GFP$^+$ cells/ DAPI$^+$ cells)$\times 10^6$. Super-resolution images were captured using a Nikon Single point scanning confocal microscope with ×40/0.95 and × 60/1.4 oil objective.

Immunofluorescence Staining for HIV-1 p24

Staining for HIV p24 protein was performed after RNAscope and before staining for nucleic acids. Cells were blocked using 1% bovine serum albumin (BSA) in 0.1% PBST (Tween20) or 5% milk in 0.1% PBST at RT for 30 min followed by antibody (Ab) labeling using FITC tagged anti-p24 antibody (10 μg/ml, Biolegend, United States) or Ms anti-HIV-1 p24 Ab (1 μg/mL, Abcam, United Kingdom) at RT for 1 h/over-night at 4°C respectively. FITC stained samples were washed twice in PBST at RT for 10 min with agitation. For secondary Ab labeling samples were washed twice in PBST at RT for 5 min on agitation prior labeling with anti-Ms IgG H&L Alexa 488 (Abcam, United Kingdom).

Specificity of RNAflow and RNAscope

Reactivated J-Lat 10.6 cells were mixed with parental Jurkat cells in 10-fold dilutions from 10 to 0.001% with the addition of 20%. Samples were then analyzed using RNAflow or RNAscope. The percentage of J-Lat 10.6 cells, determined by dilution scale, were used for prediction of NPM for each sample in the dilution series. Detected values were compared to the predicted values to determine the linearity and Limit of Detection (LOD) of the assay. LOD was calculated with the corresponding linear equation and $y = mean_{blank} + 3SD_{blank}$. All the experiments were performed at least three technical replicates with three biological replicates.

Specificity of RNAflow and RNAscope were tested using three controls; negative control (J.Lat 10.6$^{Activated-/Probe+}$), technical negative control (J.Lat 10.6$^{Activated+/Probe-}$) and positive control (J.Lat 10.6$^{Activated+/Probe+}$).

RNAscope to Detect Diverse HIV-1 Subtypes RNA

To evaluate the specificity of RNAscope to HIV-1 group M subtypes, specific clones (A1, B, C, 01_AE and 02_AG) were isolated from stored plasma samples of randomly chosen adult patients at Karolinska University Hospital, Stockholm, Sweden, and used for transfection of HEK 293T cells. The pNL4-3 plasmid was used as a positive control (NIH AIDS reagent program, United States). The subtypes were confirmed to be pure (HIV-1A1, HIV-1B, and HIV-1C) or circulating recombinant form (01_AE and 02_AG) by near full-length genome sequencing (Aralaguppe et al., 2016).

Cloning of Patient-Derived HIV-1 Gag-Pol

Viral RNA was extracted from 140 μL plasma using QIAmp Viral RNA Mini Kit (Qiagen, Germany). For a generation of the complementary strand, superscript III RT enzyme (Invitrogen, United States) was used with the gene-specific primer 6231R, as described by our group (Grossmann et al., 2015). The high-fidelity KAPA HiFiHotStart Ready Mix (KAPA Biosystem, United States) was used for two rounds of PCR. The first-round primers were 0682F and 6231R. Second-round primers 0702F-*BssHII* (GCGCGCCTAGAAGGAGAGAGAGATGGGTGCGAG) and 5798R-*SalI* (GTCGACCTCTCATTGCCACTGTCTTCTGCTC) contained the restrictions sites *BssHII* and *SalI* (New England Biolabs, United States) respectively that were subsequently used for cloning in the pNL4-3Δ*gag-pol* plasmid and validated by sequencing as described recently (Neogi et al., 2018).

Transfection With Clones From HIV-1 Subtypes

Patient-derived clones were transfected in HEK 293T cells in a 1:3 dilution of DNA: Fugene (Promega, United States) using 1 μg DNA for 24 h. All transfection reagents were mixed in Opti-MEM^{TM} (Gibco, Life Technologies, Carlsbad, United States) to a proportion of 1:10 of total media volume of RPMI. After 24 h post-transfection, media was removed and cells washed twice in PBS prior fixation using 10% NBF for 20 min at RT. Subsequently, RNAscope was performed and IF using primary and secondary antibody staining. Images were acquired using single point scanning confocal microscopy.

PBMCs Isolation and Detection

As a proof-of-concept, EDTA-blood samples were obtained from treatment naïve HIV-1 infected patients with viremia (n = 2) as well as HIV-1 negative individuals (n = 2). Patient 1 (PT#01) was infected with CRF01_AE and had a plasma viremia of 1970000 copies/mL while patient 2 (PT#02) was infected with HIV-1C with a plasma viremia of 69700 copies/mL. PBMCs were isolated by using Ficoll-Paque Plus reagent (GE Healthcare Life Sciences, Piscataway, NJ) according to the manufacturer's protocol. Cells were frozen in 10% dimethyl sulfoxide (DMSO)–90% FBS after isolation. Upon preparation, donor PBMCs were allowed to recover 24 hrs in RPMI (10% FBS, 0.2% Penstrep and 20 units IL-2 (1 μg/mL, PeproTech). After recovery, LRA activation was conducted prior RNAscope procedure, including IF staining using FITC tagged Ab, as described above, to determine the NPM value. The experiments were performed two technical replicates with two biological replicates.

Ethical Considerations

The study was approved by regional ethics committees of Stockholm (2013/1944–31/4). All participants gave informed consent, and patients identity was anonymized and delinked before analysis.

RESULTS

There are currently no reliable markers for a precise detection and quantification of latently infected cells *ex vivo* nor *in vitro*. Studies aiming at quantifying the proportion of transcription and translation competent latently infected cells through biomolecule detection are dependent on transcriptional activation of the provirus. We explored the capacity of mRNA detection *in situ* to detect HIV-1 transcripts by flow cytometry (RNAflow) and microscopy (RNAscope) to evaluate the sensitivity and specificity of each detection system following latency reversal using LRAs (**Figure 1**).

LRAs Optimization

In our study, the optimal combination of LRAs for reactivation of latent HIV-1 was Prostatin (6 μM) and hTNFα (10 ng/mL), as measured by GFP expression, and cell viability (**Figure 2**). In J-Lat6.3 and J-Lat9.2, hTNFα + Prostatin (20 and 38.2% respectively) hade a significant increase in activation ($p < 0.05$) compared to hTNFα/SAHA (0.6 and 3.5% respectively), hTNFα/Ionomycin (1.5 and 6.2% respectively), Prostatin/SAHA (6.18 and 13.3% respectively) and Prostatin/Ionomycin (10.3 and 21% respectively) while in J-Lat10.6 hTNFα + Prostatin (88%) hade a more significant increase in activation than hTNFα/SAHA (51%) and hTNFα/Ionomycin (66.7%). This combination LRA (cLRA) was therefore used for all experiments.

Optimization of RNAscope Workflow to Improve Cell Recovery Rate

Extensive washing steps tend to affect the overall output of both RNAflow and RNAscope. For RNAflow we observed ∼70–82% of cell loss while performing all assays in the low-binding Eppendorf tube, while RNAscope procedure contributed to a ∼70–78% loss of cells on coverslips following the centrifugation in the CytoSpin *centrifuge* (Thermo Fisher, United States). To improve cell coverage, we aimed at optimizing the coating strategy of coverslips to gain sufficient electrostatic interactions between cell membrane (negatively charged ions) and coverslips (positively charged ions) using Poly-L-lysine, without changing the 3D morphology. Poly-L-Lysine pre-treatment positively affected the cell coverage shown by an increase in the attachment of activated J-Lat 10.6 cells, as measured by GFP (**Supplementary Figure 1A**). The optimization induced an increase in cell retention from 22.4 to 85.9%, from the input number of cells after the complete RNAscope procedure (**Supplementary Figure 1B**). Poly-L-Lysine is a strong coating agent that mediates an increase in positively charged sites available for binding on the coated surface thereby retaining cells during multiple washing steps.

Specificity and Limit of Detection (LOD) of RNAflow in Latently Infected Cells

The quantified proportion of activated HIV$^{mRNA+/GFP+}$ cells using PrimeFlow^{TM} was 77% of total cells compared to 0.043 and 0.14% detection in the negative and technical controls, respectively. HIV$^{mRNA-/GFP-}$ cells were detected at a rate of 21.9, 99.6, and 30.2% in J-Lat 10.6$^{Activation+/Probe+}$, J-Lat

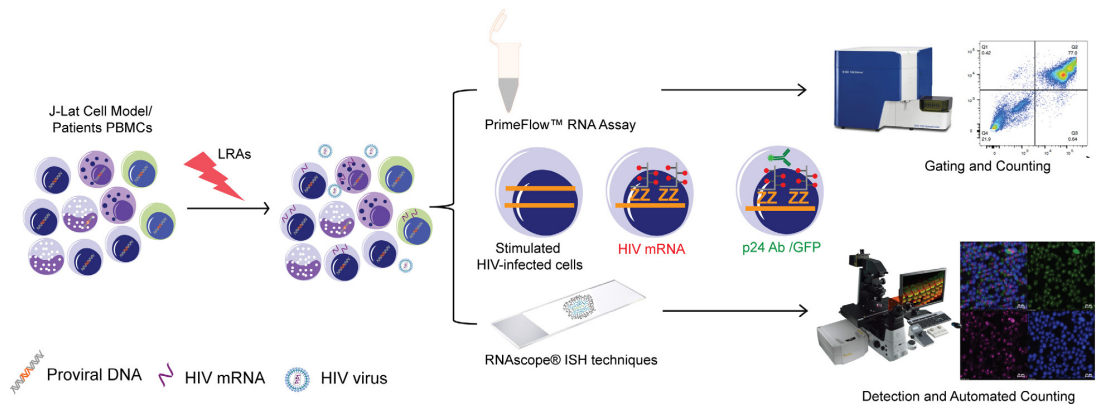


FIGURE 1 | Study designing and experimental plan: Schematic representation of experimental plan for RNAflow and RNAscope.

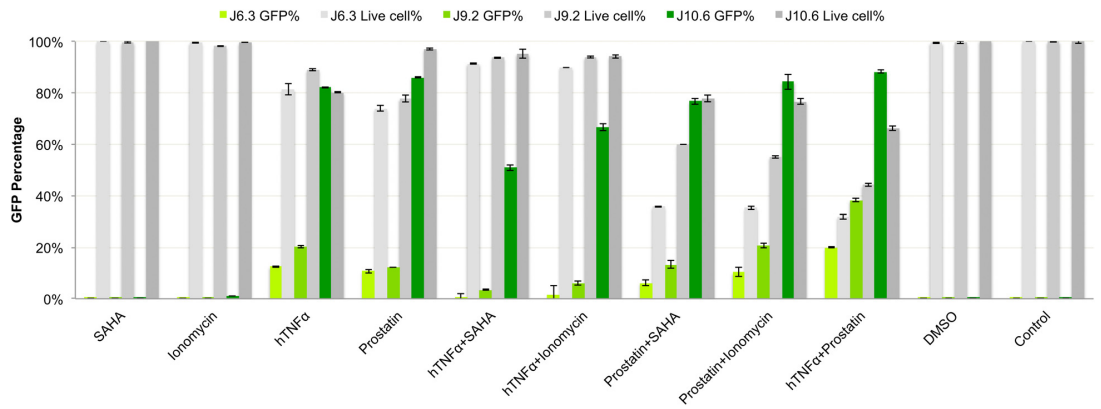


FIGURE 2 | Optimization of latency reversing reagents (LRAs). Three J-Lat cell models 6.3, 9.2, and 10.6 were used to screen for optimal LRA combinations. The doses used were: hTNFα: 10 ng/ml, Prostratin: 6 μM, Ionomycin: 1.25 μM, SAHA: 6.25 μM as a single agent or in combinations. Percentage of GFP positive cells and viability were measured. J-Lat 10.6 showed maximum activation and was further used for RNA detection using the RNAflow and RNAscope. The LRA combinations hTNF-α and prostatin showed maximum activation in all three-cell model and further used as cLRAs.

$10.6^{\text{Activation}-/\text{Probe}+}$ and J-Lat $10.6^{\text{Activation}+/\text{Probe}-}$, respectively (**Figure 3A**). The LOD of the PrimeFlow assay was determined to be 575 normalized cell numbers/million (NPM) (**Figure 3B**). One of the major disadvantages of using RNAflow is the high cell loss during the preparational stages. We intended to analyze 1×10^6 cells. Following RNAflow, we were able to acquire 1×10^5 to 2×10^5 cells with an input of 1×10^6 cells due to cell loss. The protocol required more than 15-fold centrifugation steps with a reduction in cell numbers for each wash. Thus, as latently infected cells can be <100 cells in one million the starting material would have to be extremely large to facilitate any reliable data. Due to the limitations in accessibility of patient material and low sensitivity, PrimeFlow™ was not an advantageous option for this application.

Specificity and Limit of Detection (LOD) of RNAscope in Latently Infected Cells

Detection of rare cells by microscopy have been made available by RNAscope due to its high sensitivity and specificity as probes and reagents are optimized to yield a low background. Here $\text{HIV}^{\text{mRNA}+/\text{GFP}+}$ could easily be identified after stimulation with LRAs using confocal microscopy (**Figure 4A**). Co-localization of the protein and mRNA signals showed a coherent detection in J-Lat $10.6^{\text{Activation}+/\text{Probe}+}$ after latency reversal whereas little or no signal was detected in J-Lat $10.6^{\text{Activation}-/\text{Probe}+}$ and J-Lat $10.6^{\text{Activation}+/\text{Probe}-}$. Using dilution series of activated J-Lat 10.6 cells in the parental Jurkat cell line, LOD was measured by automatic cell counting and analysis for dual recognition of $\text{HIV}^{\text{mRNA}+/\text{GFP}+}$. LOD for RNAscope was significantly reduced to 45 NPM compared to RNAflow (**Figure 4B**). The NIS element software allowed for automated wide-field analysis and quantification of the proportion of cells with on-going viral replication, thereby determined as reactivation of latency (**Supplementary Figure 2**). Automated detection removed any bias incorporated due to human errors and thus contributed to a highly, unbiased detection system as all mRNA signal are related to cells that have an active transcription of GFP, and total number of cells based on detected nuclei following DAPI staining.

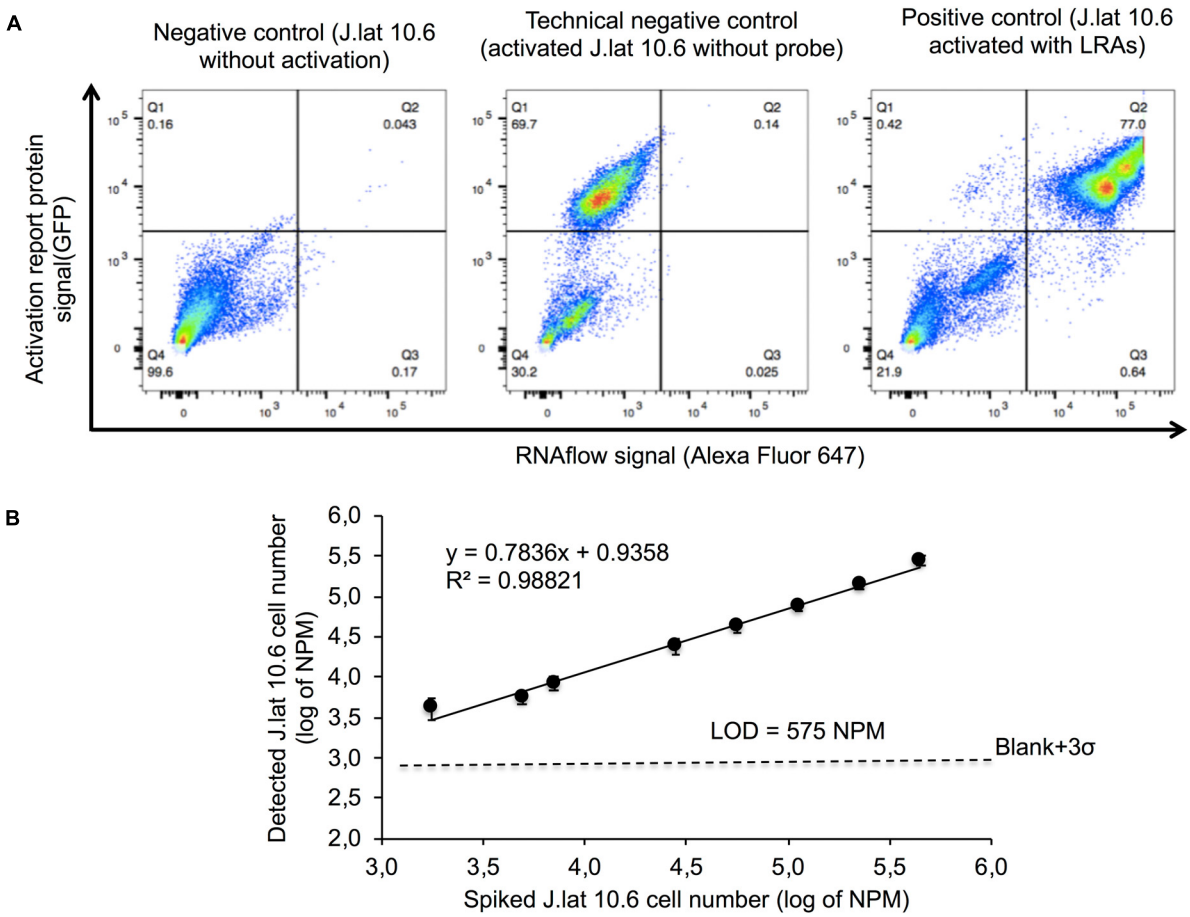


FIGURE 3 | Gating strategy and determination of the lower detection limit of RNAflow. Reactivated J-Lat 10.6 cells were spiked into parent Jurkat cells in 10-fold dilutions from 10% to 0.001% with the addition of 20% population. **(A)** The gating strategy for RNA positive cells (lower right panel), GFP and RNA positive cells (upper right panel) and GFP positive cells (upper left panel). Specificity of RNAflow were tested using three controls; negative control (J.Lat 10.6$^{\text{Activated}-/\text{Probe}+}$), technical negative control (J.Lat 10.6$^{\text{Activated}+/\text{Probe}-}$) and positive control (J.Lat 10.6$^{\text{Activated}+/\text{Probe}+}$). **(B)** Detected values were compared to the predicted values to determine the linearity and Limit of Detection (LOD) of the assay which was calculated using the corresponding linear equation ($y = \text{mean}_{\text{blank}} + 3\text{SD}_{\text{blank}}$).

RNAscope to Detect Diverse HIV-1 Subtypes

Within the HIV-1 genome, the *gag-pol* region is known to be structurally, and evolutionary semiconserved (Li et al., 2015). HIV-1 has a high mutational rate, and thereby large variations occur on a genetic level between the subtypes. The original probe set was designed for HIV-1B subtype. To investigate if these probes could detect viral mRNA from various HIV-1 subtypes, HEK 293T cells were transfected with pNL4-3*gag-pol* clones of HIV-1A1, HIV-1B, HIV-1C, and 01_AE isolated from patients. The HIV$^{\text{RNA}+}$ detection was complemented with immunofluorescence (IF) detection of the viral protein p24. This protein is translated early on during HIV-1 replication thereby making it a good validating marker for activation of latency. HIV$^{\text{RNA}+/\text{p24}+}$ dual detection allowed for combinational validation of each cell harboring translation-competent virus. The probes were capable of detecting all subtypes using a standardized transfection for 24 h (**Figure 5**). Thus, these probes target a region that is semiconserved within the HIV-1 genome, indicating the capability of utilizing this method to quantify the proportion of latently infected cells in patient material independent on subtype.

Detection of Translation-Competent Latently Infected Cells *ex vivo* Using RNAscope

In vitro models never wholly convey the inherent complexity of a biological system. Therefore, we aimed at evaluating HIV-1 reactivation from latency in clinical samples *ex vivo* using PBMCs from two therapy-naïve individuals. In the absence of latency reversal 319 and 54 cells/million PBMCs were HIV$^{\text{mRNA}+/\text{p24}+}$ in Patient#1 and Patient#2, respectively, after normalizing the background with an HIV-negative control using RNAscope (**Figure 6A**). After LRAs stimulation, the number of detectable HIV$^{\text{mRNA}+/\text{p24}+}$ cells increased to 710 and 184 in Patient#1 and Patient#2,

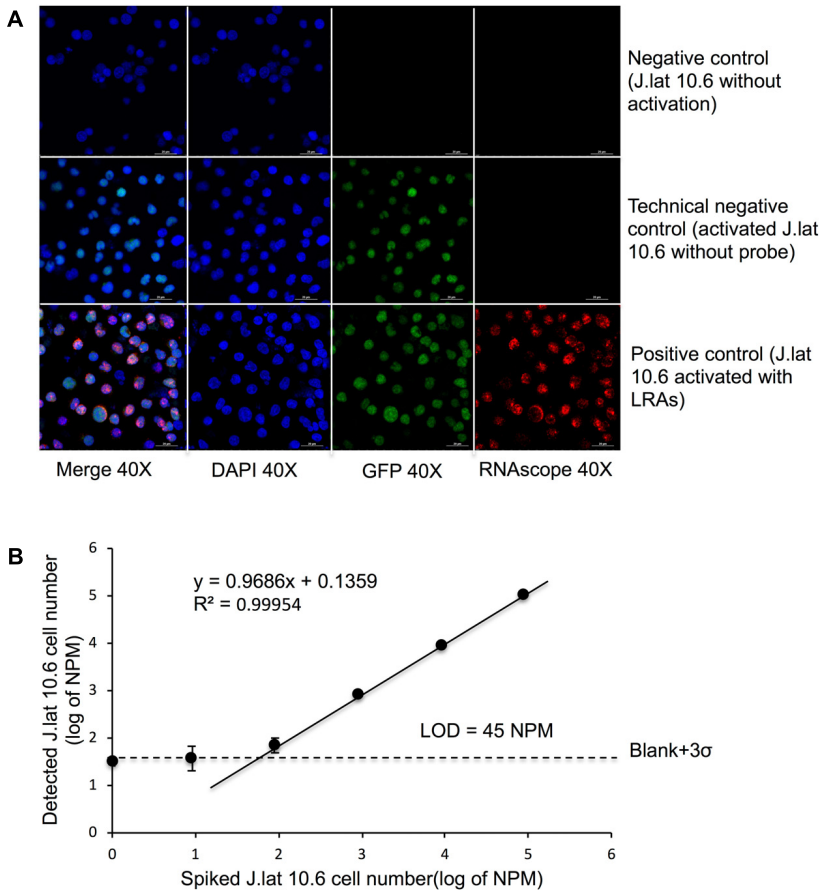


FIGURE 4 | Automated counting strategy and the lower detection limit of RNAscope **(A)** Detection of rare cells by microscopy following RNAscope HIV$^{mRNA+/GFP+}$, HIV^{mRNA+} and HIV^{GFP+} cells were detected after LRA stimulation using confocal microscopy with a series (as mentioned in **Figure 3**) of activated J-Lat 10.6 cells in the parental Jurkat cell line **(B)** LOD was measured by automatic cell counting and analysis for dual recognition of HIV$^{mRNA+/GFP+}$ using the same formula mentioned in **Figure 3**. Automated counting detected all mRNA signal related to cells that have an active transcription of GFP, and correlates it to the proportion of all detected nuclei (DAPI).

respectively, indicating a reactivation of latent virus within the PBMC (**Figure 6A**). The HIV^{mRNA+} cells were found more frequently than dual (HIV$^{mRNA+/p24+}$) positive cells but in a lower frequency than HIV^{p24+}. In one patient-sample HIV^{p24+} cells were detected in a more substantial amount than HIV^{mRNA+} cells, but a higher standard deviation was found compared to HIV^{mRNA+} cells. As individual patients are known to have different proportions of latently infected cells, detectable HIV mRNA varied to a high degree between our two patients. A high percentage of viral mRNA could be caused by a large number of latently infected cells in therapy naïve individuals or by a high efficacy of the LRAs promoting ongoing viral replication. By wide-filed analysis, all HIV$^{mRNA+/p24+}$ positive cells could be detected as single cells exhibiting viral mRNA correlating with anti-p24 detection (**Figure 6B**) indicating the applicability of the method at single-cell level. The capacity for detecting HIV$^{mRNA+/p24+}$ cells in clinical samples indicates a translational potential of this method from *in vitro* to *ex vivo* applications.

DISCUSSION

In this study, the RNAflow and RNAscope techniques were evaluated using LRAs to induce the HIV-1 latent reservoir in a small volume of input cells. Our modified RNAscope performed better that RNAflow due to a limited cells loss and a lower detection limit of 45 cells/million cells. Although the probes for RNAscope were developed for HIV-1B, they enabled the detection of several other patient-derived HIV-1 subtypes (HIV-1A1, HIV-1B, HIV-1C, 01_AE). The method can also allow single cell characterization of HIV-1 RNA and protein simultaneously.

The fluorescence *in situ* hybridization-flow cytometry technique have been reported to be sensitive and specific in determining HIV-1 mRNA and protein after latency reactivation at the single-cell level. (Martrus et al., 2016; Baxter et al., 2017; Grau-Exposito et al., 2017). Earlier, the RNAflow studies have reported to use 1 x 10^6 cells (Martrus et al., 2016). However the 18 washing steps needed contributed to a substantial loss of cells (∼70–82%), which is in line with previous studies which observed nearly a 70% cell loss (Baxter et al., 2017;

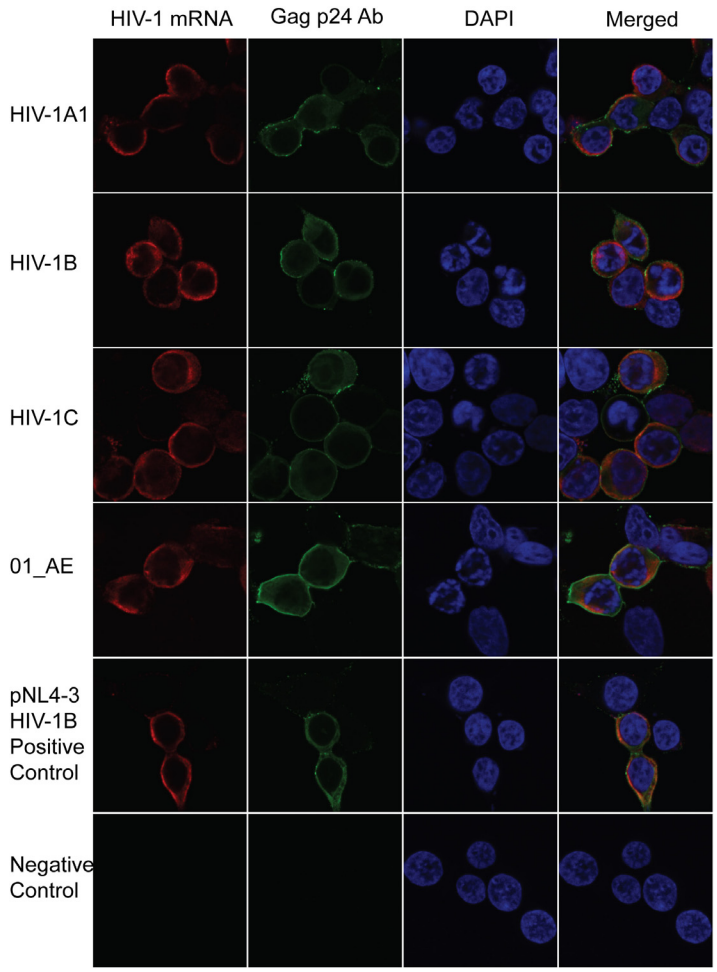


FIGURE 5 | Subtype specificity of RNAscope using patients derived chimeric viruses of HIV-1 subtypes. Chimeric viruses having HIV-1A1, B, C and CRF 01_AE gag-pol recombinant plasmids were transfected in 293T cells. HIV-1 mRNA and p24 Gag protein was detected 24 h post-transfection by RNAscope and antibody staining.

Grau-Exposito et al., 2017). Therefore, final cells for analysis ended up to be around 2×10^5 to 3×10^5 from 1×10^6 cells as starting material. This large proportion of cell loss is an important factor for the compromised lower detection limit for RNAflow which makes it in the present form of limited value to study the reservoir *ex vivo*. In contrast, we managed to get a higher sensitivity for the RNAScope. A recent study showed simultaneous detection of multiple viral nucleic acid intermediates and proteins using RNAscope at a single-cell resolution (Puray-Chavez et al., 2017). Here we evaluated the same protocol but with surface modification of the coverslips for non-adherent J-Lat cells and PBMCs which reduced the cell loss to less than 20%. Therefore we were able to scan nearly 8×10^5 in the assay which contributed to a lower detection limit of 45 cells/million cells, 10-fold lower than RNAflow.

Among the several challenges of HIV-1 cure research, high-throughput drug screenings are essential to identify new promising LRAs efficiently. Also, the effect of combinations of LRAs or compounds that target multiple pathways controlling latency can be explored to identify the viral re-activation process (Rasmussen and Lewin, 2016). However, studies on *ex vivo* latency reversal, of LRAs targeting different cellular pathways, have given contradicting results. Often LRAs reactivate HIV-1 nonuniformly across different cells models (Spina et al., 2013) and frequently two drugs in combination increase the viral reactivation. E.g., a combination of PMA and Ionomycin increased the viral reactivation *ex vivo* when compared to PMA or Ionomycin alone (Spina et al., 2013; Baxter et al., 2017; Grau-Exposito et al., 2017). However, in our study we excluded PMA since it has been shown that PMA induction of HIV-1 replication can be Tat-independent (Luznik et al., 1995) and false results can be obtained when using PMC (Spina et al., 2013). An earlier study indicated that PKC agonists reactivated latent HIV-1 uniformly across different cell models (Spina et al., 2013). Our data showed that the PKC agonstists hTNF-α and prostatin gave slightly better reactivation in all three cell lines. The combination was also able to reactivate the viral reservoir in patients PBMCs.

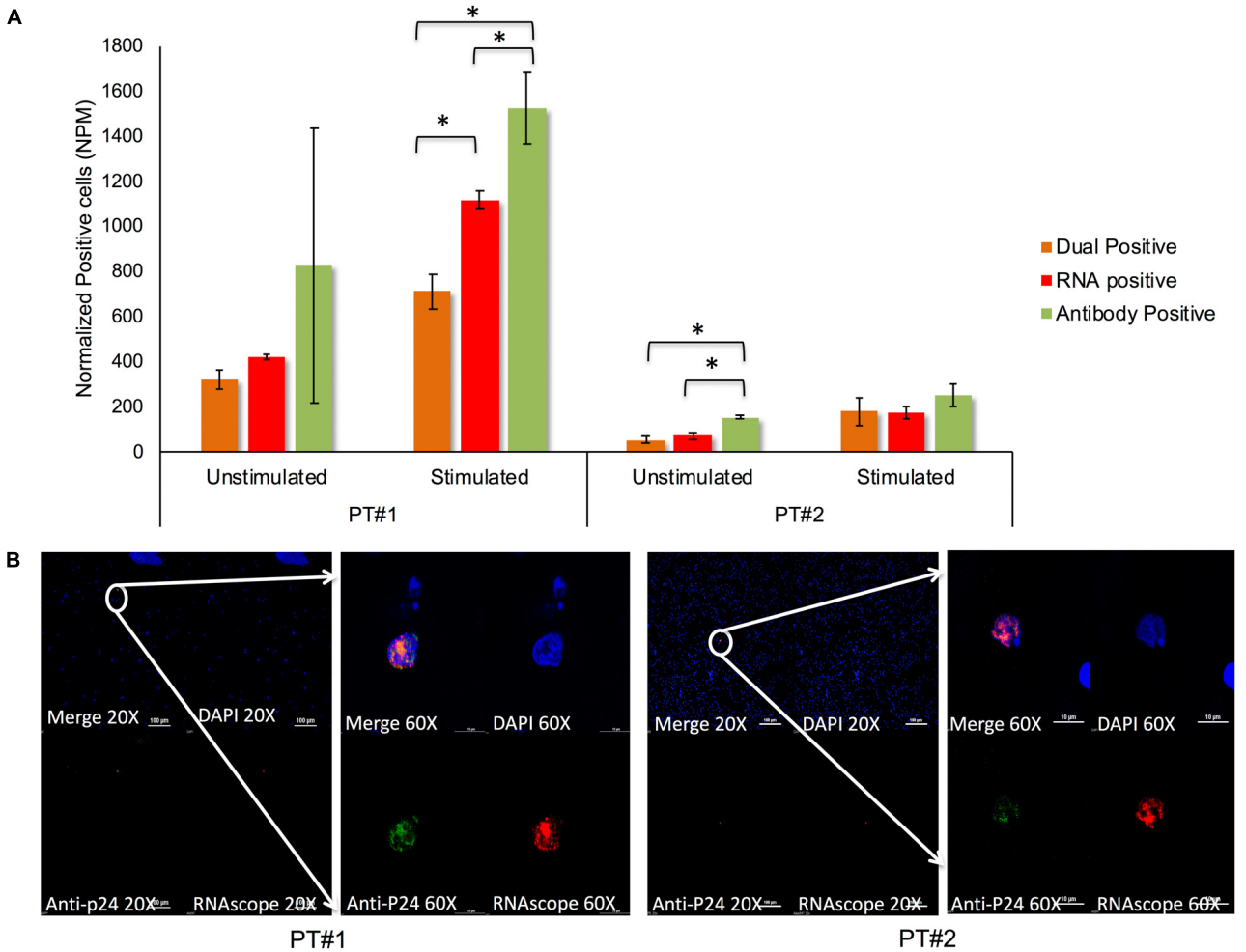


FIGURE 6 | Counting of translation-competent cells in patients derived PBMC by RNAscope. **(A)** *Ex vivo* detection of HIV-1 mRNA and p24 protein in clinical samples using PBMCs from two therapy-naïve individuals with and without cLRA activation. After cLRAs stimulation, detectable $HIV^{mRNA+/p24+}$ cells increased to 710 and 184 in Patient#1, Patient#2 indicating a reactivation of latent virus within the clinical samples. Patient 1 was infected with CRF01_AE viruses and had plasma viremia 1970000 copies/mL while patient 2 was infected with HIV-1C viruses with plasma viremia 69700 copies/mL. $p < 0.05$ are marked with '*' and consider significant. **(B)** Wide-field analysis, all $HIV^{mRNA+/p24+}$ positive cells could be detected as single cells exhibiting viral mRNA correlating with anti-p24 detection indicating the applicability of the method at single-cell level.

Another challenge in HIV-1 molecular assays is the genetic variation of the subtypes. Both RNAflow and RNAscope probe sets were developed for subtype B. To the best of our knowledge, earlier studies have not commented on the subtype specificities of these assays (Martrus et al., 2016; Baxter et al., 2017; Grau-Exposito et al., 2017; Puray-Chavez et al., 2017). We tested the RNAscope HIV-1B probe sets for patient-derived HIV-1A1, HIV-1B, HIV-1C, and 01_AE subtypes and were able to detect the viral mRNA for all the subtypes.

Our study has limitations that merit comments. First, RNAscope can detect as low as 45 cells/million, but the standard deviation near the lower detection limit is also high in the patient samples tested. However, more sensitivity can be achieved by scanning a large number of cells. Since the number of remaining latently HIV-1 infected cells are low after long-term successful cART (Joos et al., 2008), it is important to further improve the lower LOD. Second, though the RNAscope probe can detect all of our subtypes, we did not test for all subtypes circulating globally. Finally, the assay can identify transcription and translation competent provirus but may not indeed represent the replication-competent latent HIV-1 reservoir as we do not measure infectious virus. However, this assay can be combined with TZM-bl cell-based assay as described (Sanyal et al., 2017).

In summary, our study showed that RNAscope could be used to quantify the number of directly observed individual cells expressing HIV-1 mRNA following cLRA. However, method needs improvement in lower detection limit to be applicable for the reservoir quantification on patients with suppressive therapy. Therefore, it can be a useful tool for characterization of translation-competent HIV-1 latently infected cells at single-cell resolution in the fields of HIV-1 pathogenesis and viral

persistence. This method can also be adapted for single-cell transcriptomics and proteomics studies.

DATA AVAILABILITY STATEMENT

All data supporting the findings in this paper are included in the main text and Supplementary Information. All relevant data are available from the authors upon request.

AUTHOR CONTRIBUTIONS

UN and WZ designed the study and experimental plan. WZ performed RNAflow and RNAscope standardization. SSA performed the RNAscope analysis with patients' derived materials. WZ and SSA wrote the method and result section. UN wrote the introduction and discussion part of the manuscript. AS provided intellectual input, the clinical material, revised the manuscript with critical input. All authors approved the final version of the manuscript.

FUNDING

The study was funded by grants from Swedish Research Council (2017-01330), and Jeanssons Stiftelser (JS2016–0185) to UN, the Swedish Research Council grant (2016-01675) to AS and ALF-grant from Stockholm County Council (ALF 20160074) to AS and UN. WZ was partially supported by China Scholarship Council (CSC). SSA was supported by Karolinska Institutet Doctoral (KID) student funding (2016-00221).

ACKNOWLEDGMENTS

The authors acknowledge the technical assistance received from Prof. Stefan Sarafianos's group at University of Missouri, United States. The microscopic part of the study was performed in part at the Live Cell Imaging Unit/Nikon Center of Excellence, Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden, supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Research Council, and the Centre for Innovative Medicine and the Jonasson Center at the Royal Institute of Technology, Sweden. The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: J-Lat Full-Length Clone (clone 6.2, 9.2, and 10.6) from Dr. Eric Verdin, Jurkat Clone E6-1 from Dr. Arthur Weiss.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02358/full#supplementary-material

REFERENCES

Aralaguppe, S. G., Siddik, A. B., Manickam, A., Ambikan, A. T., Kumar, M. M., Fernandes, S. J., et al. (2016). Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. *J. Virol. Methods* 236, 98–104. doi: 10.1016/j.jviromet.2016.07.010

Baxter, A. E., Niessl, J., Fromentin, R., Richard, J., Porichis, F., Massanella, M., et al. (2017). Multiparametric characterization of rare HIV-infected cells using an RNA-flow FISH technique. *Nat. Protoc.* 12, 2029–2049. doi: 10.1038/nprot.2017.079

Bruner, K. M., Hosmane, N. N., and Siliciano, R. F. (2015). Towards an HIV-1 cure: measuring the latent reservoir. *Trends Microbiol.* 23, 192–203. doi: 10.1016/j.tim.2015.01.013

Churchill, M. J., Deeks, S. G., Margolis, D. M., Siliciano, R. F., and Swanstrom, R. (2016). HIV reservoirs: what, where and how to target them. *Nat. Rev. Microbiol.* 14, 55–60. doi: 10.1038/nrmicro.2015.5

Eriksson, S., Graf, E. H., Dahl, V., Strain, M. C., Yukl, S. A., Lysenko, E. S., et al. (2013). Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog.* 9:e1003174. doi: 10.1371/journal.ppat.1003174

Grau-Exposito, J., Serra-Peinado, C., Miguel, L., Navarro, J., Curran, A., Burgos, J., et al. (2017). A novel single-cell fish-flow assay identifies effector memory CD4(+) T cells as a major niche for HIV-1 transcription in HIV-infected patients. *mBio* 8:e00876-17. doi: 10.1128/mBio.00876-17

Grossmann, S., Nowak, P., and Neogi, U. (2015). Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management. *J. Int. AIDS Soc.* 18:20035. doi: 10.7448/IAS.18.1.20035

Ho, Y. C., Shan, L., Hosmane, N. N., Wang, J., Laskey, S. B., Rosenbloom, D. I., et al. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155, 540–551. doi: 10.1016/j.cell.2013.09.020

Joos, B., Fischer, M., Kuster, H., Pillai, S. K., Wong, J. K., Boni, J., et al. (2008). HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16725–16730. doi: 10.1073/pnas.0804192105

Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* 22, 1868–1877. doi: 10.1093/emboj/cdg188

Karlsson, A. C., Lindkvist, A., Lindback, S., Gaines, H., and Sonnerborg, A. (2001). Recent origin of human immunodeficiency virus type 1 variants in resting CD4 + T lymphocytes in untreated and suboptimally treated subjects. *J. Infect. Dis.* 184, 1392–1401. doi: 10.1086/324405

Laird, G. M., Eisele, E. E., Rabi, S. A., Lai, J., Chioma, S., Blankson, J. N., et al. (2013). Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* 9:e1003398. doi: 10.1371/journal.ppat.1003398

Li, G., Piampongsant, S., Faria, N. R., Voet, A., Pineda-Peña, A.-C., Khouri, R., et al. (2015). An integrated map of HIV genome-wide variation from a population perspective. *Retrovirology* 12:18. doi: 10.1186/s12977-015-0148-6

Luznik, L., Kraus, G., Guatelli, J., Richman, D., and Wong-Staal, F. (1995). Tat-independent replication of human immunodeficiency viruses. *J. Clin. Invest.* 95, 328–332. doi: 10.1172/JCI117660

Martrus, G., Niehrs, A., Cornelis, R., Rechtien, A., Garcia-Beltran, W., Lutgehetmann, M., et al. (2016). Kinetics of HIV-1 latency reversal quantified on the single-cell level using a novel flow-based technique. *J. Virol.* 90, 9018–9028. doi: 10.1128/JVI.01448-16

Neogi, U., Singh, K., Aralaguppe, S. G., Rogers, L. C., Njenda, D. T., Sarafianos, S. G., et al. (2018). Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes. *AIDS* 32, 469–476.

Procopio, F. A., Fromentin, R., Kulpa, D. A., Brehm, J. H., Bebin, A. G., Strain, M. C., et al. (2015). A novel assay to measure the magnitude of the inducible viral reservoir in HIV-infected individuals. *eBio Med.* 2, 872–881. doi: 10.1016/j.ebiom.2015.06.019

Puray-Chavez, M., Tedbury, P. R., Huber, A. D., Ukah, O. B., Yapo, V., Liu, D., et al. (2017). Multiplex single-cell visualization of nucleic acids and protein during HIV infection. *Nat. Commun.* 8:1882. doi: 10.1038/s41467-017-01693-z

Rasmussen, T. A., and Lewin, S. R. (2016). Shocking HIV out of hiding: where are we with clinical trials of latency reversing agents? *Curr. Opin. HIV AIDS* 11, 394–401. doi: 10.1097/COH.0000000000000279

Sanyal, A., Mailliard, R. B., Rinaldo, C. R., Ratner, D., Ding, M., Chen, Y., et al. (2017). Novel assay reveals a large, inducible, replication-competent HIV-1 reservoir in resting CD4(+) T cells. *Nat. Med.* 23, 885–889. doi: 10.1038/nm.4347

Spina, C. A., Anderson, J., Archin, N. M., Bosque, A., Chan, J., Famiglietti, M., et al. (2013). An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4 + T cells from aviremic patients. *PLoS Pathog.* 9:e1003834. doi: 10.1371/journal.ppat.1003834

Weiss, A., Wiskocil, R. L., and Stobo, J. D. (1984). The role of T3 surface molecules in the activation of human T cells: a two-stimulus requirement for IL 2 production reflects events occurring at a pre-translational level. *J. Immunol.* 133, 123–128.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 December 2018
doi: 10.3389/fmicb.2018.02799

Combining Phylogenetic and Network Approaches to Identify HIV-1 Transmission Links in San Mateo County, California

Sudeb C. Dalai[1,2†], Dennis Maletich Junqueira[3,4*†], Eduan Wilkinson[3,4], Renee Mehra[5], Sergei L. Kosakovsky Pond[6], Vivian Levy[1,7], Dennis Israelski[1,7], Tulio de Oliveira[3,4,8] and David Katzenstein[1]*

[1] Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States, [2] Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA, United States, [3] KwaZulu-Natal Research Innovation and Sequencing Platform, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, [4] School of Laboratory Medicine and Medical Science, Department of Health Sciences, University of KwaZulu-Natal, Durban, South Africa, [5] Division of Hematology, Stanford University School of Medicine, Stanford, CA, United States, [6] Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States, [7] San Mateo Medical Center, San Mateo, CA, United States, [8] Department of Global Health, University of Washington, Seattle, WA, United States

OPEN ACCESS

Edited by:
Michael M. Thomson, Instituto de Salud Carlos III, Spain

Reviewed by:
Juan Angel Patino Galindo, Columbia University, United States
Samuel R. Friedman, National Development and Research Institutes, United States

***Correspondence:**
Sudeb C. Dalai
sdalai@stanford.edu;
sudeb.dalai@gmail.com
Dennis Maletich Junqueira
dennismaletich@hotmail.com

† These authors are joint first authors

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *07 August 2018*
Accepted: *31 October 2018*
Published: *06 December 2018*

Citation:
Dalai SC, Junqueira DM, Wilkinson E, Mehra R, Kosakovsky Pond SL, Levy V, Israelski D, de Oliveira T and Katzenstein D (2018) Combining Phylogenetic and Network Approaches to Identify HIV-1 Transmission Links in San Mateo County, California. Front. Microbiol. 9:2799. doi: 10.3389/fmicb.2018.02799

The HIV epidemic in San Mateo County is sustained by multiple overlapping risk groups and is an important hub for HIV transmission in northern California. Limited access to care has led historically to delayed clinical presentation, higher rates of opportunistic infections, and an increased prevalence of antiretroviral drug resistance. The virologic and clinical consequences of treatment within these multiple ethnic and behavioral groups are poorly understood, highlighting the need for efficient surveillance strategies that are able to elucidate transmission networks and drug resistance patterns. We obtained sequence data from a group of 316 HIV-positive individuals in the San Mateo AIDS Program over a 14-year period and integrated epidemiologic, phylogenetic, and network approaches to characterize transmission clusters, risk factors and drug resistance. Drug resistance mutations were identified using the Stanford HIV Drug Resistance Database. A maximum likelihood tree was inferred in RAxML and subjected to clustering analysis in Cluster Picker. Network analysis using pairwise genetic distances was performed in HIV-TRACE. Participants were primarily male (60%), white Hispanics and non-Hispanics (32%) and African American (20.6%). The most frequent behavior risk factor was male-male sex (33.5%), followed by heterosexual (23.4%) and injection drug use (9.5%). Nearly all sequences were subtype B (96%) with subtypes A, C, and CRF01_AE also observed. Sequences from 65% of participants had at least one drug resistance mutation. Clustered transmissions included a higher number of women when compared to non-clustered individuals and were more likely to include heterosexual or people who inject drugs (PWID). Detailed analysis of the largest network (N = 47) suggested that PWID played a central role in overall transmission

of HIV-1 as well as bridging men who have sex with men (MSM) transmission with heterosexual/PWID among primarily African American men. Combined phylogenetic and network analysis of HIV sequence data identified several overlapping risk factors in the epidemic, including MSM, heterosexual and PWID transmission with a disproportionate impact on African Americans and a high prevalence of drug resistance.

Keywords: HIV, transmission links, California, phylogenetics, network

INTRODUCTION

Molecular epidemiologic analyses of the HIV-1 epidemics in Africa (Dalai et al., 2009; Gray et al., 2009; Mir et al., 2016), Asia (Neogi et al., 2012; Pang et al., 2012), and North America (Smith et al., 2010; Mehta et al., 2015) have provided evidence for distinct epidemic dynamics and patterns of transmission within defined communities. While the southern African and Asian epidemics are thought to be sustained by sexual-social factors, high-risk people who inject drugs (PWID), and commercial sex work (CSW) (Thorne et al., 2010; Baral et al., 2012; de Oliveira et al., 2017), male-male sex has remained the primary mode of transmission in the United States, accounting for approximately 70% of new infections (Centers for Disease Control and Prevention, 2015). In the past 10 years, the HIV epidemic in California has shifted from a primarily white MSM (men who have sex with men) population to include a diverse range of overlapping risk groups, where heterosexual women comprise the fastest growing demographic for new infections and an increasing fraction of new diagnoses occur among diverse racial and ethnic groups (California Department of Public Health, 2015). Hispanic monolingual men and women now predominate among new HIV-1 cases in several California counties followed by African Americans and Asian/Pacific Islanders (Magis-Rodríguez et al., 2009).

The HIV epidemic in San Mateo County (California) encompasses approximately 1,613 diagnosed cases with 0.2% adult prevalence[1]. The epidemic is sustained by multiple ethnic, migratory, and behavioral networks, including MSM, migrant populations from Asia and Latin America, and PWID, each having distinct patterns of HIV acquisition and transmission (San Mateo County Department of Public Health, 2017). Within these communities, access to care, including antiretroviral treatment (ART) has historically been constrained by financial, cultural, and linguistic barriers. HIV-positive immigrants (78.7% Hispanic) attending the publicly-funded San Mateo County AIDS Program have substantially delayed clinical presentation, marked by lower baseline CD4+ cell counts, greater prevalence of opportunistic infections, and higher hospitalization rates as compared with United States-born individuals (Levy et al., 2006). Moreover, population-based surveys in San Mateo have recently identified multiple risk behaviors associated with immigrant status, including unprotected sex, unstable and overlapping sexual partnerships, CSW contacts, and PWID (Levy et al., 2005). The implications of prolonged, undiagnosed HIV infection, viremia, delayed treatment, and expanded risk behavior for community transmission networks remain largely uncharacterized.

In California, a high prevalence of drug resistance among untreated, newly diagnosed patients has highlighted widespread community-level transmission and cross-border introduction of multi-drug resistant HIV associated with an expansion in migratory and risk behaviors (Panichsillapakit et al., 2016). In recent years a focus on addressing health disparities in California has increased access to treatment and care for historically marginalized populations. However, data are limited regarding the virologic and clinical outcomes of treatment in these communities. These observations underscore the need for broad and systematic surveillance of HIV transmission networks and drug resistance in the context of comprehensive health service delivery systems, particularly among populations who present late and where ART treatment access is constrained.

Defining the patterns of HIV transmission and resistance within communities is important for regional prevention and treatment programs to develop effective, integrated testing and treatment strategies to reduce transmission and to identify and appropriately treat newly-infected individuals. This study combines phylogenetic inference, network analysis and molecular virology/epidemiology to characterize HIV viral transmission and ARV drug resistance in a northern California community epidemic over a 14-year period.

MATERIALS AND METHODS

Study Population

The study population included 316 HIV-positive adults receiving care as part of the publicly funded San Mateo County AIDS Program who underwent clinically indicated genotypic antiretroviral resistance testing (GART) from 1996 to 2010. The population included both acutely- and chronically-infected, as well as treatment naïve and multi-drug experienced individuals. Demographic (age, gender, race/ethnicity), epidemiologic (date of diagnosis, mode of transmission, social/risk behavior, partner information, location data), and clinical information (history of ART usage, HIV clinical stage, HIV viral load, CD4+ cell count, co-infections) were de-identified and extracted from electronic and written medical records at San Mateo Medical Center. Mode of HIV transmission was extracted from medical records as determined by the physician at the time of patient intake. Transmission categories included (1) MSM; (2) PWID; (3) MSM + PWID; and (4) heterosexual/other (including participants reporting infection through contaminated blood

[1] http://www.countyhealthrankings.org/

products). Missing values were treated as a separate category for all demographic variables analyzed in this study.

The use of anonymized, de-identified clinical/demographic and sequence data was reviewed and approved under an exempt protocol by the Institutional Review Boards of Stanford University, the University of California, Berkeley, and Mills-Peninsula Health Services on behalf of San Mateo Medical Center. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Sequence Data and Alignment

Genotypic resistance assays were performed as part of standard clinical care by the Stanford Hospital Clinical Virology Laboratory. A total of 637 HIV-1 *pol* gene sequences were obtained from 316 patients tested over the study period. Sequences were generated by dideoxynucleotide sequencing of population (consensus) amplicons from HIV-1 *pol*. Sequences were aligned and manually edited using the ClustalW algorithm as implemented in BioEdit (Hall, 1999). For individuals who had multiple sequences, the earliest available sequence was retained for transmission/clustering analyses as well as for the screening of drug resistance mutations.

Additionally, a reference sequence dataset was compiled using BLAST+ (Camacho et al., 2009). For each study sequence, the 50 most similar reference sequences from BLAST were selected, resulting in a reference alignment of 1,405 sequences after removal of duplicates.

Subtype Classification and Drug Resistance Analysis

HIV-1 subtype and evidence for inter-subtype recombination were assessed using the REGA Subtyping Tool v3.0 (de Oliveira et al., 2005; Alcantara et al., 2009). The HIVseq algorithm was used to interpret genotypic resistance and to identify known ART drug resistance mutations (DRMs) according to the most recent International AIDS Society (IAS) mutation list (Wensing et al., 2017). Both methods were implemented in the Stanford HIV Drug Resistance Database (Gifford et al., 2009).

Phylogenetic and Clustering Analysis

A maximum likelihood (ML) phylogenetic tree was constructed in RAxML (Stamatakis, 2014) using the general time reversible model of nucleotide substitution (Tavaré, 1986), an estimated proportion of invariant sites and a gamma correction for among-site rate variation. Statistical support for internal nodes was obtained via bootstrapping with 1000 replicates. Alternatively, branch supports were also calculated via transfer bootstrap expectation (TBE) in BOOSTER (Lemoine et al., 2018) using all 1000 replicates generated in RAxML. TBE provides optimized support for deep branches in large phylogenetic trees when compared to the classical bootstrap criterion, allowing for the identification of large putative clusters. Trees were visualized using FigTree v1.4.3 (Rambaut, 2009).

In order to assess potential bias from DRMs, comparative phylogenetic trees were constructed in PhyML (Guindon et al., 2005) from a separate sequence dataset including only samples from San Mateo. All IAS codons associated with major antiretroviral drug resistance (Wensing et al., 2017) were excluded from one of these alignments. Reliability of the obtained topologies was estimated with 1000 bootstrap replicates. Notably, inclusion or removal of DRMs did not appreciably alter tree structure.

Transmission clusters were identified with the program Cluster Picker (Ragonnet-Cronin et al., 2013) using a minimum branch support of 90 (bootstrap or BTE) and an intra-cluster genetic distance threshold of 4 percent or 8 percent. Clusters including more than 5 individuals were separately evaluated in TempEst to investigate their temporal signal. When a molecular clock assumption was validated, these clusters were submitted for Bayesian phylogenetic reconstruction using BEAST v1.8.4 (Rambaut et al., 2016; Suchard et al., 2018). In addition to a phylogenetic approach, HIV-TRACE was used to reconstruct a putative genetic transmission network with a pairwise genetic distance of 2%, which is in the 1 – 2% range derived from several comparative studies of epidemiologically-linked partners and studies of within-host evolutionary rates (Wertheim et al., 2014). HIV-TRACE performs agglomerative hierarchical clustering connecting sequences only if their pairwise distance does not exceed the assigned threshold. As an analog to phylogenetic bootstrap, i.e., to assess network sensitivity to sampling error, we repeated network inference on 100 bootstrap replicates created from the original alignment with goalign[2]. Medical records were independently reviewed to identify epidemiologic linkages among study participants irrespective of the phylogenetic/network linkages. Univariate and multivariate statistical analyses of risk factors associated with clustering were performed in the statistical package R (R Core Team, 2013).

RESULTS

Patient Demographics

Between 1996 and 2010, 637 HIV sequences were obtained from 316 people living with HIV/AIDS (PLWHA) receiving ART in the San Mateo County AIDS Program. **Table 1** summarizes baseline demographic characteristics of the participants. The study population was comprised primarily of white Hispanics and non-Hispanics (32.0%) and African Americans (20.6%). Around 60% of participants were male, with ages ranging from 19 to 67 years old. The most frequent recorded mode of HIV transmission was MSM (33.5%), followed by heterosexual (23.4%), PWID (9.5%), and combined MSM/PWID (8.5%). Three participants reported infection through contaminated blood products.

Transmission Clustering

Of the 316 genotypes, nearly all sequences were subtype B (95.9%, **Figure 1**). Six sequences (1.9%) were subtype C, four individuals (1.3%) were infected with the circulating recombinant form CRF01_AE, and three sequences (0.9%) were subtype A.

Sequences from San Mateo were combined with 1,405 reference sequences identified as the most similar sequences from

[2]https://github.com/fredericlemoine/goalign

TABLE 1 | Demographic and epidemiological data for 316 HIV-positive adults receiving antiretroviral treatment in San Mateo County, California.

	Clustered Sequences		Total (*n*=316)
	Cluster Picker (*n*=66)	HIV-TRACE (*n*=61)	
Gender			
Male	32 (48.48%)	31 (50.82%)	190 (60.13%)
Female	25 (37.88%)*	18 (29.51%)*	49 (15.5%)
Transgender/Unknown	9 (13.64%)	12 (19.67%)	77 (24.37%)
Age (years)	45.2 (1.1)	43.8 (1.2)	41.7 (0.5)
Primary Language			
English	47 (71.21%)	39 (63.93%)	185 (58.5%)
Spanish	5 (7.58%)	4 (6.56%)	26 (8.2%)
Bilingual	3 (4.55%)	2 (3.28%)	12 (3.8%)
Other/Unknown	11 (16.67%)	16 (26.23%)	93 (29.4%)
Ethnicity			
Black	39 (59.09%)*	23 (37.7%)*	65 (20.6%)
White	7 (10.61%)	14 (22.95%)	101 (32.0%)
Asian	1 (1.52%)	4 (6.56%)	13 (4.1%)
Latino	8 (12.12%)	6 (9.84%)	46 (14.6%)
Other/Unknown	11 (16.67%)	14 (22.95%)	91 (28.8%)
Hispanic ancestry			
No	48 (72.73%)*	40 (65.57%)	187 (59.2%)
Yes	9 (13.64%)	9 (14.75%)	53 (16.8%)
Unknown	9 (13.64%)	12 (19.67%)	76 (24.1%)
Mode of transmission			
Heterosexual	29 (43.94%)*	23 (37.7%)*	74 (23.4%)
PWID	19 (28.79%)*	12 (19.67%)*	30 (9.5%)
MSM	4 (6.06%)	11 (18.03%)	106 (33.5%)
MSM/PWID	5 (7.58%)	2 (3.28%)	27 (8.5%)
Unknown	9 (13.64%)	13 (21.31%)	79 (25.0%)
Drug Resistance Mutation			
PI	17 (25.76%)	8 (13.11%)	87 (27.5%)
NRTI	38 (57.58%)	22 (36.07%)	186 (58.9%)
NNRTI	11 (16.67%)	4 (6.56%)	74 (23.4%)

*Data are n (%) or mean (SE). MSM, men who have sex with men; PWID, people who inject drugs, injection drug users; PI, protease inhibitor; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non- nucleoside reverse transcriptase inhibitors. *$p < 0.05$ (cluster vs. non-clustered comparisons for each method).*

GenBank. Cluster Picker identified 8 and 9 phylogenetic clusters (TBE bootstrap > 90%) at intra-cluster genetic thresholds of 4 and 8%, respectively. At the 4% intra cluster genetic distance threshold, the phylogenetic tree included seven transmission pairs and one putative cluster of three individuals. At an 8% cut-off, eight transmission pairs and one large single cluster (n = 50) were identified. No significant correlation between sampling date and evolutionary rate was found for this large cluster in TempEst (slope: −0.0013, R2: 0.093443), which implies that phylodynamic characterization would not be indicated. Nevertheless, sequences were submitted to a Bayesian analysis and after 4×10^8 million steps, runs did not converge (ESS > 13) and precluded further estimates. Network analysis with HIV-TRACE detected seven transmission pairs and a single putative transmission network of 47 individuals (**Figure 2**). Five of the seven transmission pairs were also identified in the Cluster Picker analysis (**Supplementary Table S1**).

Cluster Picker and HIV-TRACE identified clusters having a similar demographic composition (**Table 1**). In a univariate analyses, clustered individuals included a higher proportion of women when compared to non-clustered patients ($p < 0.001$). In addition, clusters were more likely to include heterosexual or PWID ($p < 0.001$) mode of transmission and African Americans ($p < 0.01$). Notably, individuals reporting MSM mode of transmission were less likely to be in clusters ($p < 0.001$). No association was seen between clustering and age or presence of DRMs (PI, NRTI or NNRTI).

Further analysis of the large network identified by HIV-TRACE suggested that PWID and MSM played a central role in the transmission of HIV-1 in this putative transmission network. Based on bootstrap support the network is divided into two different sub-clusters (**Figure 2**): (i) the larger subgroup includes 27 PWID or heterosexual individuals that are linked with a high, well-supported degree of connectedness (2.7 links per person supported by a bootstrap > 70%) and has a overall genetic diversity of 0.022 [bootstrap procedure standard error (SE): 0.002062], (ii) a second cluster bridging to the first by two weakly supported links (63 and 65% bootstraps) is comprised mainly of MSM individuals (n = 9) followed by heterosexual (n = 5) and unknown risk group individuals (n = 6) and has an intra-genetic diversity of 0.027 (SE: 0.002315). The genetic distance between sub-groups was 0.033 (SE: 0.003280).

Drug Resistance

Of 305 participants with available drug resistance data, 65% had at least one DRM. The most frequent nucleoside reverse transcriptase inhibitor (NRTI) resistance mutations were M184V (49.5%), T215Y (26.9%), and M41L (20.3%). The most frequent non-nucleoside reverse transcriptase inhibitor (NNRTI) mutations were K103N (15.4%), G190E (3%), and K101E (2.6%). The most frequent protease inhibitor (PI) mutations were at positions 46 (8.9%), 54 (7.2%) and 82 (26%). Factors associated with drug resistance included: male gender; chronic HIV infection; diagnosis with HIV and genotyping performed during an earlier (as opposed to more recent) time period; lower viral load (3.9 vs. 4.4 $\log_{10}$ copies/mL, $p = 0.006$); and NRTI or NNRTI treatment. After multivariate adjustment, factors independently associated with drug resistance of any class included being male, genotyping during an earlier time period, lower viral load, and exposure to NNRTIs.

DISCUSSION

Phylogenetic analyses have been widely used to define and characterize transmission links among HIV-infected individuals (Baral et al., 2012; Wertheim et al., 2014; Rambaut et al., 2016; de Oliveira et al., 2017; Kostaki et al., 2018). San Mateo County has a unique and heterogeneous population of individuals living with HIV/AIDS with distinct ethnic, racial and language communities overlapping the traditional behavioral risk-groups of MSM, heterosexual and PWID. Relative to county demographics,

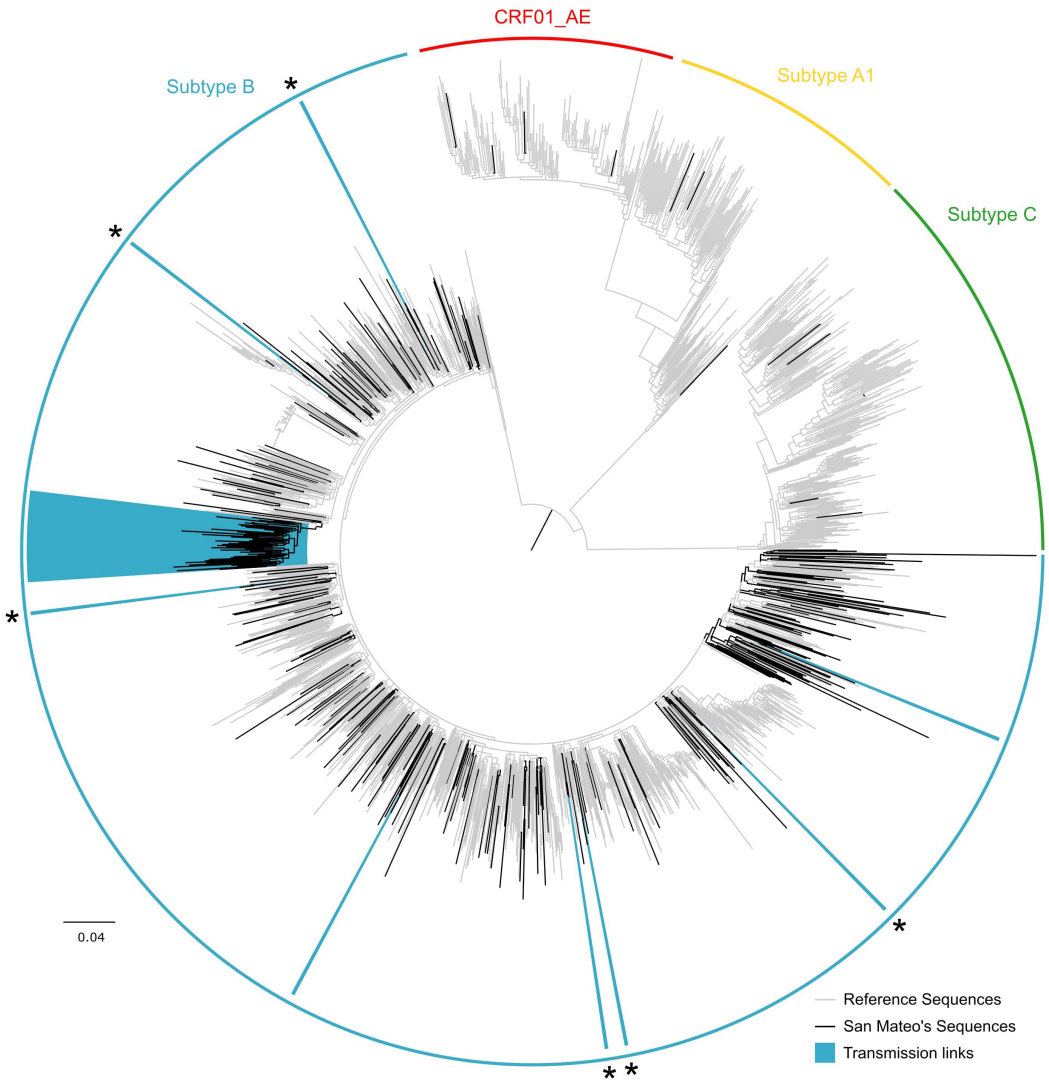


FIGURE 1 | Maximum likelihood HIV-1 *pol* transmission clusters in San Mateo County, California. Colored semi-circles indicate HIV subtypes. Black branches represent sequences obtained from 316 HIV-positive adults receiving antiretroviral treatment as part of the publicly funded San Mateo County AIDS Program. Gray branches denote reference sequences. Clades highlighted in blue indicate transmission clusters identified in Cluster Picker (TBE ≥ 90 and genetic distance threshold of 8%). Asterisks indicate transmission pairs mutually identified as transmission links in Cluster Picker and HIV-TRACE. The scale bar at the bottom represents the number of substitutions per site along branches in the tree topology.

African American individuals are over-represented and Asian Americans underrepresented among patients presenting for evaluation and treatment in the San Mateo AIDS Program (San Mateo County Sexually Transmitted Disease and HIV-AIDS Surveillance Annual Report, 2015).

Combined phylogenetic and network analyses among HIV-positive adults from San Mateo County have identified a striking role of intra-community transmission dynamics among African Americans and highlight the importance of epidemiological bridging between risk groups in the local epidemic. As expected, we found some differences in the number and composition of clusters detected using two different methods (Rose et al., 2017), which may be attributed to different underlying methodologies. HIV-TRACE uses pairwise genetic distances between isolates to build putative transmission networks, while Cluster Picker relies on the intra-cluster genetic distance within putative clusters in a phylogenetic tree in conjunction with branch support. By using a minimum branch support of 90% (TBE) and an intra-cluster genetic distance threshold of 8%, we identified nine supported clusters in Cluster Picker. In contrast, with a maximum pairwise genetic distance of 2%, the transmission network approach in HIV-TRACE detected eight clusters (**Supplementary Table S1**). Transmission pairs had a concordance of 67% between the two methods. Interestingly, both methods identified one large cluster of linked individuals (**Figures 1**, **2**). In addition, the clusters commonly identified had similar demographic composition.

Network analysis of the large 47-person cluster using HIV-TRACE suggests the existence of a highly-connected linked chain

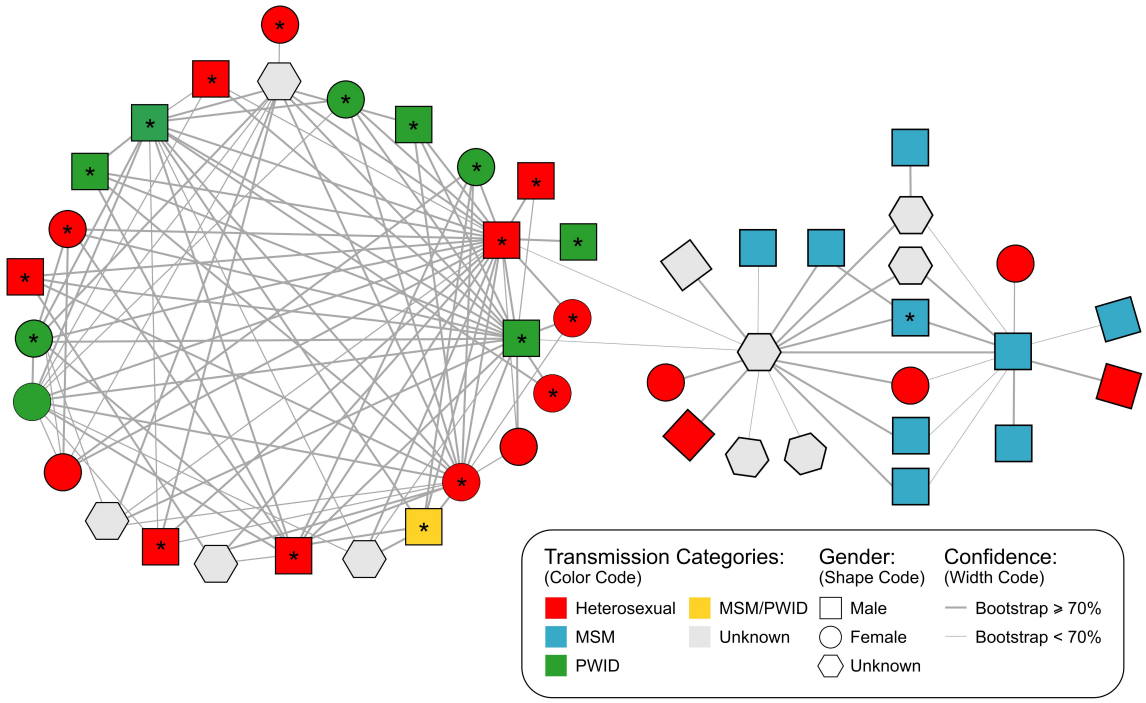


FIGURE 2 | Inferred genetic network structure of a major cluster (N = 47 individuals) in the San Mateo HIV-1 subtype B epidemic identified by HIV-TRACE using a genetic distance threshold of 2%. Shape codes gender, colors denote the reported mode of transmission for each individual and line width represents the bootstrap confidence for linkage. *are indicating African-American individuals. MSM, men who have sex with men; PWID, people who inject drugs, injection drug users.

of transmission comprising individuals of diverse demographic backgrounds. This network included two sub-clusters with different primary modes of HIV transmission (**Figure 2**). Twenty-seven individuals were substantially interconnected in a heterosexual and/or PWID transmission chain mainly involving African American individuals. Connected to this cluster is a less-intricate network primarily composed of white MSM. The linkage between the two sub-clusters included two males (PWID and heterosexual) and an individual of unknown gender but highly connected to MSM. In such a scenario, some individuals may provide a bridge between the MSM and heterosexual epidemics mostly likely through PWID. Alternatively, this may reflect the well-documented epidemiological phenomena of bisexual black males who do not identify as gay and do not disclose high-risk MSM behavior to their female partners, but subsequently contribute disproportionately to heterosexual HIV transmission (Millett et al., 2005; Bond et al., 2009). The fact that the inter-genetic distance between sub-clusters is higher than that within groups may explain the weak association (63 and 65% bootstraps) between the two sub-clusters and may potentially reflect the result of a remote transmission link.

Risk factors associated with transmission clustering include heterosexual or PWID transmission modes, female gender and African descent. MSM was not associated with clustering in this sample. These results are consistent with epidemiologic observations in San Mateo County in San Mateo County where a diverse range of overlapping risk groups, particularly heterosexual women, is identified among new infections (Levy et al., 2006). A previous study of 96 large United States metropolitan areas demonstrated higher AIDS incidence and mortality among heterosexuals in areas with a higher population prevalence of HIV-positive PWID as well as decreased presence of needle-exchange or drug-use treatment programs (Friedman et al., 2014). These findings corroborate our results and suggest that the burden of HIV/AIDS among heterosexuals in more recent years may have been shaped by bridging from HIV-infected PWID to heterosexuals and possibly, though to a lesser degree, from HIV-positive MSM to heterosexuals. These results highlight that efforts to mitigate transmission among PWID, including HIV counseling and testing or drug use treatment, can directly impact transmission to other key populations (World Health Organization, 2012; Friedman et al., 2014). Despite the predominance of Hispanic monolingual men and women among new HIV-cases in California (Levy et al., 2005), no clusters in this study had a significant representation of Spanish-speaking individuals. However, a substantial number of individuals included in clusters were diagnosed before 1999, suggesting that our sampling and methods identified remote transmission links that may not necessarily reflect current epidemic dynamics.

This study has some limitations and potential biases. The demographic composition of patients in the San Mateo AIDS Program differs substantially from the overall population in San Mateo County; notably, Black or African Americans accounted for 20.6% of the study population while comprising only 3% of the county population annually, and Asians accounted for only 4.1% of the study population while annually they comprise

24–29% of the county demographic. Patients attending the San Mateo AIDS Program also have lower average income than the county median (Levy et al., 2006), highlighting the health inequalities underlying transmission of HIV and other sexually transmitted infections.

Finally, the sampling frame defined within San Mateo County is restrictive. Several study subjects had previously moved from neighboring areas to San Mateo to access treatment, whereas others indicated they were infected or diagnosed in other US states or outside the United States. In comparable community-based studies comprised of predominantly MSM, large transmission clusters have been identified (Smith et al., 2010; Mehta et al., 2015). The lack of large clustering in the present study despite nearly 33% MSM participants suggests the existence of uncaptured, overlapping transmission networks in other geographic locations. For example, the lack of clustering among monolingual White Hispanics may reflect transitory migration and infection with HIV strains that originated along the Mexico-California border, or, alternatively, a bias toward selecting older strains that were circulating earlier in the epidemic, as immigrant status has been independently associated with delayed presentation, diagnosis and care (Sanchez et al., 2004; Levy et al., 2005, 2006; Dennis et al., 2015). Finally, there was incomplete availability of epidemiological and clinical data largely due to destruction or remote storage of archived medical records, limiting the statistical power of clustering and drug resistance analyses.

CONCLUSION

As HIV genotypic data become routinely available for molecular epidemiologic analyses, initial treatment options and public health approaches to ART implementation can be optimized to avoid early virologic failure particularly in the setting of resource limitations. The high frequency of drug resistance (65.5%) identified in this population most likely reflects a substantial burden of long-standing HIV disease with heavy ART exposure (89.2%). This highlights the importance of genotypic testing to determine the most effective ARV regimen for newly-diagnosed individuals. Our finding of over 4% non-subtype B HIV in the San Mateo AIDS Program (largely among immigrants and transitory individuals), including subtypes A, C, and recombinant CRF01_AE, provides evidence for the introduction of HIV variants other than subtype B into the community. Integration of phylogenetic and network methodologies to identify putative transmission links detected a large cluster with demographic and risk group composition reflecting the HIV epidemic in San Mateo County. Female gender, African-descent and heterosexual intercourse or PWID seem to be key features of these transmission networks. Despite aggressive epidemiologic surveillance, HIV awareness campaigns, and behavior change programs, the rate of new HIV infections has increased or remained constant in many United States communities. High-resolution molecular epidemiologic analyses are useful means of identifying sources of viral diversity and excess transmission risk and, when integrated with broad surveillance, can contribute to strategies to reduce new HIV infections.

DATA AVAILABILITY

GenBank accession numbers of the sequences described in this study: MK025232 – MK025548.

AUTHOR CONTRIBUTIONS

SD, TdO, and DK conceived and designed the research. SD, DJ, EW, RM, VL, and SLKP conducted the research. SD, RM, VL, and DI involved in the patient data collection and clinical care. SD, RM, and VL performed the sample collection and molecular analysis. SD, DJ, EW, and SLKP performed the data analysis and created graphics. SD and DJ wrote the manuscript, with the assistance of all co-authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02799/full#supplementary-material

REFERENCES

Alcantara, L. C. J., Cassol, S., Libin, P., Deforche, K., Pybus, O. G., Van Ranst, M., et al. (2009). A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res.* 37, 1–9. doi: 10.1093/nar/gkp455

Baral, S., Beyrer, C., Muessig, K., Poteat, T., Wirtz, A. L., Decker, M. R., et al. (2012). Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12, 538–549. doi: 10.1016/S1473-3099(12)70066-X

Bond, L., Wheeler, D. P., Millett, G. A., LaPollo, A. B., Carson, L. F., and Liau, A. (2009). Black men who have sex with men and the association of down-low identity with HIV risk behavior. *Am. J. Public Health* 99(Suppl. 1), 92–95. doi: 10.2105/AJPH.2007.127217

California Department of Public Health (2015). *California HIV Surveillance Report*. Sacramento, CA: California Department of Public Health.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Centers for Disease Control and Prevention (2015). *HIV Surveillance Report*. Atlanta, GA: Centers for Disease Control and Prevention.

Dalai, S. C., de Oliveira, T., Harkins, G. W., Kassaye, S. G., Lint, J., Manasa, J., et al. (2009). Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. *AIDS* 23, 2523–2532. doi: 10.1097/QAD.0b013e3283320ef3

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., et al. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21, 3797–3800. doi: 10.1093/bioinformatics/bti607

de Oliveira, T., Kharsany, A. B. M., Gräf, T., Cawood, C., Khanyile, D., Grobler, A., et al. (2017). Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV* 4, e41–e50. doi: 10.1016/S2352-3018(16)30186-2

Dennis, A. M., Hué, S., Pasquale, D., Napravnik, S., Sebastian, J., Miller, W. C., et al. (2015). HIV transmission patterns among immigrant Latinos illuminated by the integration of phylogenetic and migration data. *AIDS Res. Hum. Retroviruses* 31, 973–980. doi: 10.1089/AID.2015.0089

Friedman, S. R., West, B. S., Tempalski, B., Morton, C. M., Cleland, C. M., Des Jarlais, D. C., et al. (2014). Do metropolitan HIV epidemic histories and programs for people who inject drugs and men who have sex with men predict AIDS incidence and mortality among heterosexuals? *Ann. Epidemiol.* 24, 304–311. doi: 10.1016/j.annepidem.2014.01.008

Gifford, R. J., Liu, T. F., Rhee, S.-Y., Kiuchi, M., Hue, S., Pillay, D., et al. (2009). The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* 25, 1197–1198. doi: 10.1093/bioinformatics/btp134

Gray, R. R., Tatem, A. J., Lamers, S., Hou, W., Laeyendecker, O., Serwadda, D., et al. (2009). Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS* 23, F9–F17. doi: 10.1097/QAD.0b013e32832faf61

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–W559. doi: 10.1093/nar/gki352

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Kostaki, E.-G., Nikolopoulos, G. K., Pavlitina, E., Williams, L., Magiorkinis, G., Schneider, J., et al. (2018). Molecular analysis of human immunodeficiency virus type 1 (HIV-1)-infected individuals in a network-based intervention (transmission reduction intervention project): phylogenetics identify HIV-1-infected individuals with social links. *J. Infect. Dis.* 218, 707–715. doi: 10.1093/infdis/jiy239

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-0

Levy, V., Page-Shafer, K., Evans, J., Ruiz, J., Morrow, S., Reardon, J., et al. (2005). HIV-related risk behavior among hispanic immigrant men in a population-based household survey in low-income neighborhoods of northern California. *Sex. Transm. Dis.* 32, 487–490. doi: 10.1097/01.olq.0000161185.06387.94

Levy, V., Prentiss, D., Balmas, G., Chen, S., Israelski, D., Katzenstein, D., et al. (2006). Factors in the delayed HIV presentation of immigrants in northern California: implications for voluntary counseling and testing programs. *J. Immigr. Minor. Heal.* 9, 49–54. doi: 10.1007/s10903-006-9015-9

Magis-Rodríguez, C., Lemp, G., Hernandez, M. T., Sanchez, M. A., Estrada, F., and Bravo-García, E. (2009). Going North: mexican migrants and their vulnerability to HIV. *JAIDS J. Acquir. Immune Defic. Syndr.* 51, S21–S25. doi: 10.1097/QAI.0b013e3181a26433

Mehta, S. R., Wertheim, J. O., Brouwer, K. C., Wagner, K. D., Chaillon, A., Strathdee, S., et al. (2015). HIV transmission networks in the San Diego–Tijuana border region. *EBioMedicine* 2, 1456–1463. doi: 10.1016/j.ebiom.2015.07.024

Millett, G., Malebranche, D., Mason, B., and Spikes, P. (2005). Focusing "down low": bisexual black men, HIV risk and heterosexual transmission. *J. Natl. Med. Assoc.* 97, 52S–59S.

Mir, D., Jung, M., Delatorre, E., Vidal, N., Peeters, M., and Bello, G. (2016). Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. *Infect. Genet. Evol.* 46, 190–199. doi: 10.1016/j.meegid.2016.05.017

Neogi, U., Bontell, I., Shet, A., De Costa, A., Gupta, S., Diwan, V., et al. (2012). Molecular epidemiology of HIV-1 subtypes in india: origin and evolutionary history of the predominant subtype C. *PLoS One* 7:e39819. doi: 10.1371/journal.pone.0039819

Pang, W., Zhang, C., Duo, L., Zhou, Y.-H., Yao, Z.-H., Liu, F.-L., et al. (2012). Extensive and complex HIV-1 recombination between B', C and CRF01_AE among IDUs in south-east Asia. *AIDS* 26, 1121–1129. doi: 10.1097/QAD.0b013e3283522c97

Panichsillapakit, T., Smith, D. M., Wertheim, J. O., Richman, D. D., Little, S. J., and Mehta, S. R. (2016). Prevalence of Transmitted HIV Drug Resistance Among Recently Infected Persons in San Diego, CA 1996–2013. *JAIDS J. Acquir. Immune Defic. Syndr.* 71, 228–236. doi: 10.1097/QAI.0000000000000831

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ragonnet-Cronin, M., Hodcroft, E., Hué, S., Fearnhill, E., Delpech, V., Brown, A. J. L., et al. (2013). Automated analysis of phylogenetic clusters. *BMC Bioinformatics* 14:317. doi: 10.1186/1471-2105-14-317

Rambaut, A. (2009). *FigTree v1.4: Tree Figure Drawing Tool*. Available at: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Rose, R., Lamers, S. L., Dollar, J. J., Grabowski, M. K., Hodcroft, E. B., Ragonnet-Cronin, M., et al. (2017). Identifying transmission clusters with cluster picker and HIV-TRACE. *AIDS Res. Hum. Retroviruses* 33, 211–218. doi: 10.1089/aid.2016.0205

San Mateo County Department of Public Health (2017). *San Mateo County Sexually Transmitted Disease and HIV-AIDS Surveillance Annual Report*. San Francisco, CA: San Mateo County Department of Public Health.

San Mateo County Sexually Transmitted Disease and HIV-AIDS Surveillance Annual Report (2015). Available from https://www.smchealth.org/sites/main/files/file-attachments/2015_stdhiv_ar_final_12_28_16.pdf.

Sanchez, M. A., Lemp, G. F., Magis-Rodríguez, C., Bravo-García, E., Carter, S., and Ruiz, J. D. (2004). The epidemiology of HIV among Mexican migrants and recent immigrants in California and Mexico. *J. Acquir. Immune Defic. Syndr.* 37(Suppl. 4), S204–S214. doi: 10.1097/01.qai.0000141253.54217.24

Smith, R. J., Okano, J. T., Kahn, J. S., Bodine, E. N., and Blower, S. (2010). Evolutionary dynamics of complex networks of hiv drug-resistant strains: the case of San Francisco. *Science* 327, 697–701. doi: 10.1126/science.1180556

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4, 1–5. doi: 10.1093/ve/vey016

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17, 57–86.

Thorne, C., Ferencic, N., Malyuta, R., Mimica, J., and Niemiec, T. (2010). Central Asia: hotspot in the worldwide HIV epidemic. *Lancet Infect. Dis.* 10, 479–488. doi: 10.1016/S1473-3099(10)70118-3

Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., et al. (2017). 2017 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 24, 132–133.

Wertheim, J. O., Leigh Brown, A. J., Hepler, N. L., Mehta, S. R., Richman, D. D., Smith, D. M., et al. (2014). The global transmission network of HIV-1. *J. Infect. Dis.* 209, 304–313. doi: 10.1093/infdis/jit524

World Health Organization. (2012). *WHO, UNODC, UNAIDS Technical Guide: For Countries to Set Targets for Universal Access to HIV Prevention, Treatment and Care for Injecting Drug Users 2012 Revision*. Geneva: World Health Organization.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 January 2019
doi: 10.3389/fmicb.2019.00020

HIV-1 Molecular Epidemiology, Transmission Clusters and Transmitted Drug Resistance Mutations in Central Brazil

Tayana Serpa Ortiz Tanaka[1], Thaysse Ferreira Leite[2], Solange Zacalusni Freitas[3], Gabriela Alves Cesar[1], Grazielli Rocha de Rezende[4], Andrea De Siqueira Campos Lindenberg[3], Monick Lindenmeyer Guimarães[2†] and Ana Rita Coimbra Motta-Castro[1,4,5†]*

[1] Programa de Pós-graduação em Doenças Infecciosas e Parasitárias, Faculdade de Medicina, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, [2] Laboratory of AIDS and Molecular Immunology, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil, [3] Universitary Hospital Maria Aparecida Pedrossian, HUMAP, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, [4] Laboratory of Clinical Immunology, FACFAN, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, [5] Laboratory of Molecular Virology, Oswaldo Cruz Foundation, Mato Grosso do Sul, Campo Grande, Brazil

OPEN ACCESS

Edited by:
Joris Hemelaar, University of Oxford, United Kingdom

Reviewed by:
Bin Su, Capital Medical University, China
Antonio Mas, University of Castilla La Mancha, Spain

***Correspondence:**
Tayana Serpa Ortiz Tanaka
tayeto@hotmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 October 2018*
Accepted: *09 January 2019*
Published: *31 January 2019*

Citation:
Tanaka TSO, Leite TF, Freitas SZ, Cesar GA, de Rezende GR, Lindenberg ADSC, Guimarães ML and Motta-Castro ARC (2019) HIV-1 Molecular Epidemiology, Transmission Clusters and Transmitted Drug Resistance Mutations in Central Brazil. Front. Microbiol. 10:20. doi: 10.3389/fmicb.2019.00020

We aimed to characterize HIV-1 molecular epidemiology and transmission clusters among heterosexual (HET) and men who have sex with men (MSM) individuals, as well as transmitted drug resistance mutations (TDRM) in Central-Western Brazil. This cross-sectional survey was conducted among 190 antiretroviral naïve HIV-1 infected individuals. Proviral DNA was extracted, and nested PCR amplified partial *polymerase* gene (PR/RT). After sequencing, subtypes were assigned, and the sequences were analyzed for the occurrence of possible transmission networks. Calibrated Population Resistance (CPR) tool from Stanford HIV Database was used to investigate the presence of TDRM. Among 150 individuals whose samples were successfully sequenced, the most prevalent HIV-1 subtype was B, followed by recombinant forms. The occurrence of twenty transmission clusters composed by at least two sequences was verified, suggesting the existence of transmission clusters among individuals from the same or distinct sexual orientations. Intermediate level of TDRM (12%) was found in the study population, and almost half of the subjects with TDRM had more than one resistance mutation. No correlations between sexual orientation and the presence of TDRM, HIV-1 subtypes/recombinants forms were verified. Taken together, the necessity of the continuous monitoring of the TDRM to verify the importance of pre-genotyping and to delineate future strategies in primary antiretroviral therapy. Likewise, the knowledge of the HIV-1 transmission networks in Brazil would allow the implementation of effective HIV-1 prevention strategies in local settings.

Keywords: HIV, MSM, molecular epidemiology, transmitted drug resistance, transmission network

INTRODUCTION

In Latin America, it is estimated that 1.8 million people are living with human immunodeficiency virus (HIV) and/or acquired immunodeficiency syndrome (AIDS). Despite 100,000 new HIV infections having been diagnosed in 2017, the HIV incidence decreased 13.7% between 2000 and 2017 (UNAIDS, 2018). In Brazil, HIV prevalence among the general population is below 0.6% and it is estimated that AIDS cases among Brazilians reached 882,810 by June 2017 (Brasil, 2017). HIV prevalence is higher in key populations at risk, for example 17.5% in men who have sex with men (MSM) (Kerr et al., 2018). The detection rate of AIDS has been falling steadily in Brazil in recent years. However, the Central Western region showed little change in its detection rate in the last 10 years, reaching 16.7 cases per 100 thousand inhabitants in 2016 (Brasil, 2017).

Universal access to combined antiretroviral therapy (cART) in Brazil was crucial in order to increase survival and decrease AIDS-related hospitalizations in HIV-1 infected individuals (Souza Junior et al., 2011). Although, the development of drug resistance mutations is a significant obstacle to maintaining HIV-1 replication suppression and can lead to viral load increase and consequently transmission of viruses with drug resistance mutations. Therefore, transmitted drug resistance mutations (TDRM) have become an important challenge, since they have been described for all drugs used in the clinical management of HIV and as incidence and prevalence vary by region this highlights the importance of its monitoring. The prevalence of TDRM could vary according to the study population, methods and lists of resistance mutations used to calculate these rates (Booth and Geretti, 2007).

Brazil has an extensive border, covering about 15,000 km, exhibiting great socioeconomic and cultural diversity across regions. Concerning HIV-1 subtypes, subtype B is the most prevalent, followed by F1, and BF1 recombinants in most Brazilian regions (De Sa Filho et al., 2005; Pedroso et al., 2007; Machado et al., 2009; Guimarães et al., 2015), except for the Southern region, where subtype C is highly prevalent (Silva et al., 2010; de Medeiros et al., 2011; Gräf et al., 2011). However, even in the same geographic region, the HIV-1 distribution could be heterogeneous (Gräf and Pinto, 2013). In border areas, intense drug trafficking and prostitution occur; both situations may affect local epidemic dynamics. Taking these geographical and epidemiological characteristics together into consideration, the study of HIV-1 genetic diversity and transmission networks as well as drug resistance mutations in this region is relevant.

MATERIALS AND METHODS

Subjects and Study Design

We conducted a cross-sectional survey among antiretroviral naïve HIV-infected individuals recruited in Campo Grande, the capital of Mato Grosso do Sul (MS) State, from 2011 to 2014. One hundred and seventy-two individuals were enrolled at Reference Centers for Parasitic and Infectious Diseases (Freitas et al., 2014), and thirty-two were MSM recruited in a cross-sectional study (Fernandes et al., 2015). Inclusion criteria were: (a) having confirmed diagnosis for HIV-1; (b) being over 18 years old; (c) being antiretroviral naïve; (d) having signed the informed consent form in earlier surveys, which predicted storage of samples and their utilization in future research; and (e) having sample stored in sufficient quantity to perform the analyses proposed. Following these criteria, 190 individuals were selected for the subsequent analysis. This study was carried out in accordance with the recommendations of the Ethical Committee on Human Research of the Federal University of Mato Grosso do Sul, that is in accordance with the Declaration of Helsinki. The protocol was approved by under protocol number 1151451, CAAE 46185915.8.0000.0021.

Amplification of HIV-1 PR/RT Region

DNA was extracted from 200 μL of each whole blood sample by using the QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The partial *polymerase* (*pol*) gene including protease/reverse transcriptase (PR/RT) region was amplified by nested polymerase chain reaction (PCR) using combinations of primers described elsewhere (Delatorre et al., 2017). The amplified products were analyzed by electrophoresis using agarose gels (1%). Amplicons were purified using the Illustra GFX® PCR DNA and Gel Band Purification Kit (GE Healthcare, United Kingdom), following the manufacturer's recommendations. The purified DNA was sequenced using Big Dye Terminator Cycle Sequencing Ready Reaction kit v.3.1 (Applied Biosystems, CA, United States) and processed with an automated ABI 3130xl sequencer (Applied Biosystems), using Sanger's method.

Sequence Analysis

The sequences were edited in DNASTAR software and then aligned with reference sequences from Los Alamos HIV Sequence Database[1] using the Clustal W program implemented in MEGA 6.0 software (Tamura et al., 2013). All sequences are available in GenBank (accession number MF545192-MF545340). The final PR/RT alignment covered a fragment of 1261 bp, corresponding to nucleotides 2254 to 3514 relative to the HXB2 genome.

Maximum Likelihood (ML) phylogenetic was constructed with the PhyML 3.0 program using an online web server (Guindon et al., 2010). The Smart Model Selection recommended the GTR+I+G nucleotide substitution model to be used in the ML (Lefort et al., 2017). The heuristic tree search was performed using the SPR branch-swapping algorithm, and the branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test (Anisimova and Gascuel, 2006). Recombinant profiles were inferred by bootscan analyses with a sliding window of 300 bp, steps of 10 bp and Kimura-2 parameters model using SimPlot 3.5.1 software (Lole et al., 1999).

Those sequences that clustered together with high aLRT support (>0.90) in the ML tree were analyzed for the occurrence of possible transmission clusters. Therefore, such sequences were submitted to analysis using nucleotide Basic Local Alignment Search Tool (BLASTn) (Altschul et al., 1990) to recover reference

[1] https://www.hiv.lanl.gov

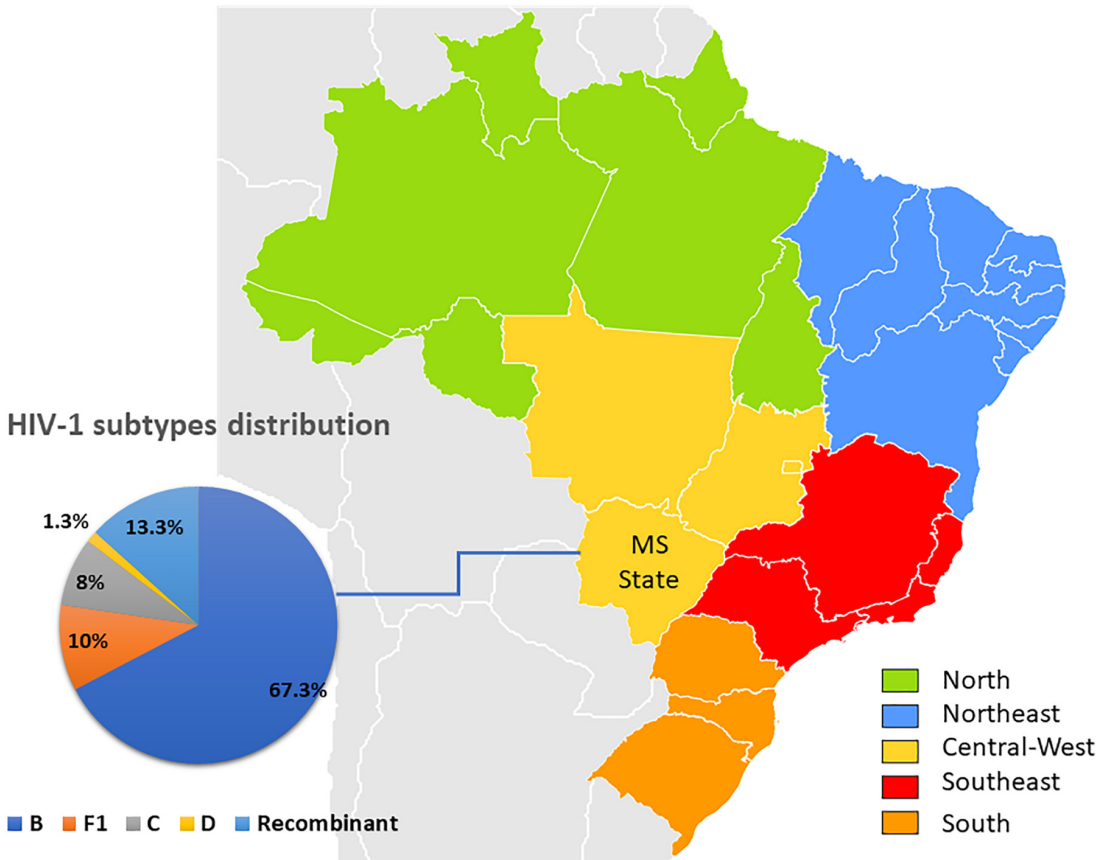


FIGURE 1 | Map of Brazil indicating Brazilian regions and MS State, from which HIV-1 sequences were obtained. The pie chart shows HIV-1 subtypes distribution based on *pol* sequences included in this study.

sequences with high similarity (>95%). These sequences retrieved were added to three new alignments from pure subtypes (B, D, and F1), and a new ML tree was obtained to verify the maintenance of the transmission clusters according to their subtypes. For subtypes D and F1 analyses we included all available Brazilian reference sequences, however, duplicate sequences were removed. For subtype B, at least ten representative sequences from each Brazilian State and all sequences from Mato Grosso do Sul state available at the Los Alamos HIV Sequence Database were included. Before performing the phylogenetic analyses to confirm the transmission clusters, drug-resistance mutations positions were stripped from each alignment, resulting in a fragment of 891 bp from nucleotides 2262 to 3251 relative to HXB2 genome. Our final cluster classification was defined based on aLRT (>90) in the phylogenetic analyses (**Figures 2**, **3**), and low mean pairwise genetic distances (≤4.5) of clustered sequences have been employed.

Genotypic Analysis of HIV-1 Drug Resistance

To investigate the presence of TDRM, the sequences were submitted to Stanford HIV Database for Transmitted DRM [TDRM/Calibrated Population Resistance Tool (CPR Tool)] Version 6.0 (Gifford et al., 2009), which uses the mutation list according to Bennett et al. (2009).

Statistical Analysis

Statistical analyses were conducted using the SPSS 17.0 statistical analysis software package (SPSS Inc., Chicago, IL, United States). Median, standard deviation (SD), range and frequencies (%) were used to describe patients' characteristics. The frequency of TDRMs was also calculated, and the chi-square or Fisher exact test was employed when appropriate. A p value of < 0.05 was defined as statistically significant.

RESULTS

Out of 190 antiretroviral naïve patients who had samples available for DNA extraction, 172 were PR/RT amplified (90.5%), and from them 150 (87.2%) were successfully sequenced. From those 150 studied subjects, 62.0% were male, with an average age of 36 years, ranging from 18 to 70 years. More than half of participants were white (53.3%), heterosexual (64.0%), and reported less than 12 years of schooling (80.7%), and irregular condom use (54%). Only 6.7% of them were sex workers.

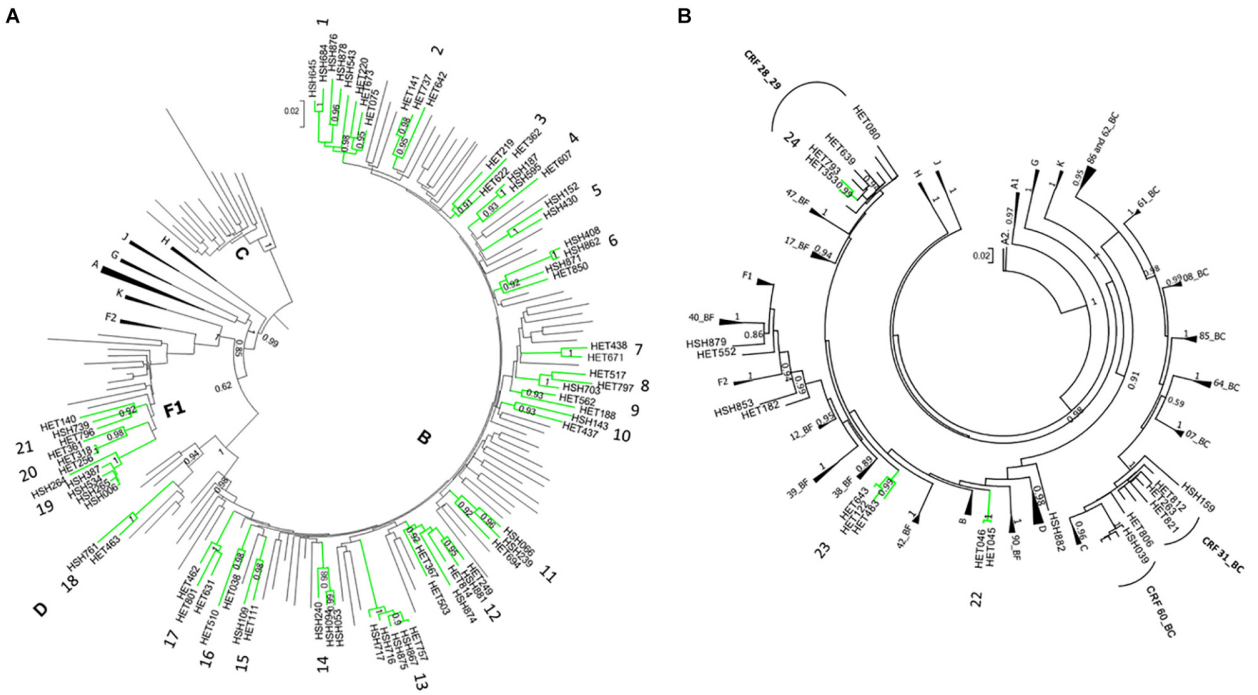


FIGURE 2 | ML phylogenetic tree of 150 HIV-1 PR/RT sequences from Mato Grosso do Sul, Central-West Brazil. The analyzed PR/RT alignment covered a fragment of 1261 bp, corresponding to nucleotides 2254 to 3514 relative to HXB2 genome. Reference sequences retrieved from GenBank are not labeled. ALRT values are represented only if greater > 0.90. Possible HIV-1 transmission clusters were indicated by numbers. HET = sample obtained from heterosexual individual; HSH = sample obtained from men who have sex with men. **(A)** Pure HIV-1 subtypes and **(B)** Recombinant sequences.

Sociodemographic and behavioral characteristics are listed in **Table 1**. No statistically significant correlation was detected between the variables presented in **Table 1** and HIV-1 subtypes.

Phylogenetical analyses revealed that 101 sequences (67.3%) were classified as subtype B, 15 (10%) as F1, 12 (8%) as C, two (1.3%) as subtype D and 20 (13.3%) possible recombinants (**Figures 1**, **2A**). The phylogenetic and bootscan analysis of these twenty sequences revealed that four (2.7%) were CRF28_29BF, three (2.0%) were CRF31_BC, one (0.7%) was CRF60_BC, and 12 (8.0%) were unique recombinant forms (URF) (**Figure 2B**).

TDRM to at least one class of antiretroviral drug was found in 18 sequences (12%), and the drug resistance mutation to nucleoside reverse transcriptase inhibitor (NRTI) was the most common (12/150; 8%), followed by non-nucleoside reverse transcriptase inhibitor (NNRTI) (7/150; 4.7%) and PI resistance (3/150; 2%) (**Table 2**). Of these, twelve (8%) were singleton mutations and six (4.0%) multiple. K103N was the most frequent resistance mutation observed (5/150; 3.3%) followed by V75M (4/150; 2.7%). There was no statistical difference between sexual orientation and the prevalence of TDRM and HIV-1 subtypes distribution.

Twenty-four possible transmission clusters, including 57 individuals were identified according to the adopted criteria (aLRT > 90 in ML analysis). The clusters involved from two to five individuals and seventeen of them belong to HIV-1 subtype B, one to subtype D, three to sub-subtype F1 (**Figure 2A**) and three were recombinant forms being 2 BF1 and 1 BD (**Figure 2B**). The inclusion of a huge number of reference sequences enabled reinvestigation by ML of the transmission clusters, in combination with the criteria of presenting high aLRT support and low mean genetic distance, allowed us to depict twenty previously identified possible transmission clusters from pure HIV-1 subtypes B, D, and F1. The possible transmission clusters 1c, 3, 10, and 21 were not confirmed. Some of the originally detected clusters remained with the same configuration (2,8,11,13,16, and 18); meanwhile, some of them presented a new shape. In the Cluster numbers (1, 4, 5, 9, 15, 17, 19, and 20) some Brazilian reference sequences clustered together to ours. We also verified that some sequences were excluded from the original possible clusters (1, 3, 6, 9, 10, 12, and 14). The possible clusters 1 and 6 give rise to two new transmission clusters (1a,b and 6a,b). The original possible clusters (**Figures 2A,B**) and the confirmed transmission clusters (**Figures 3**, **4**) were summarized in **Table 3**. Since the clusters BD (22) and BF1 (23 and 24) were unique recombinant forms, we did not perform an additional ML phylogenetic tree.

All subtype B sequences were classified as pandemic B. Among subtype B confirmed clusters, twelve (12/17; 70.6%) had more than two sequences, and five (5/17; 29.4%) were composed of two sequences. Five clusters comprised MSM samples of this study with or without other Brazilian sequences (clusters 1a, 1b, 5, 6a, and 14), four with HET samples (clusters 2, 7, 16, and 17), six, mixed HET, and MSM sequences (clusters 4, 8, 11, 12, 13, and 15). Two clusters (6b and 9) were formed by one sequence from our

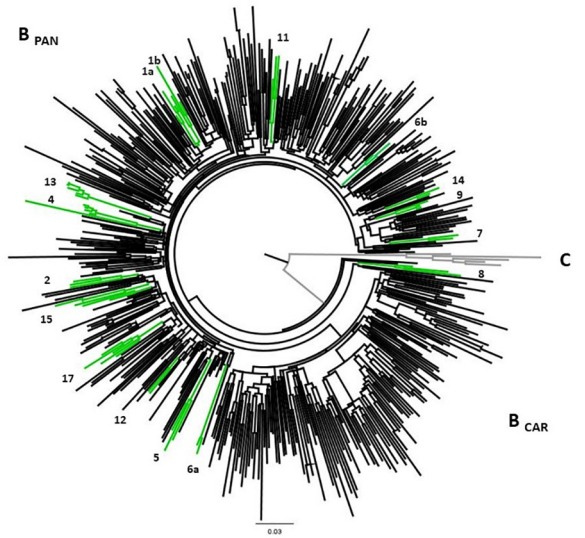


FIGURE 3 | ML phylogenetic tree highlighting HIV-1 subtype B transmission clusters. The confirmed HIV-1 transmission clusters were highlighted in green and numbered according to the previous grouping from **Figure 1**. All clusters present aLRT ≥ 0.90 and low mean pairwise genetic distances (≤4.5). The analysis involved 520 HIV-1 B PR/RT sequences (102 sequences from the present study, 331 Brazilian reference sequences, 82 non-Brazilian reference sequences, and 5 HIV-1 Subtype C sequences as outgroup). The analyzed fragment corresponds to 891 bp (2262 to 3251 nt relative to HXB2 genome) and drug-resistance mutations positions were stripped.

study and two other Brazilian sequences from MS state, retrieved from Genbank (**Table 3**).

Individuals from ten clusters of subtype B were positive for lifetime syphilis and/or Hepatitis B and C infections. Four (4/17; 23.5%) contained sequences with TDRM, and two of them (clusters 1b and 4) were composed by MSM sharing the same TDRM. Cluster 1b included two MSM who had a history of Treponema pallidum infection and K103N mutation, and one of them reported being a sex worker and bisexual. Cluster 4 grouped two sequences from MSM (HSH187 and HSH595), one from a male HET, and sequences BRMS58 and BRMS14_10, both from males (da Silveira et al., 2012), all of them had the V75M substitution, associated with resistance to NRTI inhibitors.

Samples belonging to non-B subtypes were grouped into three clusters (**Figure 4**). Two of them (19 and 20), belonging to F1 subtype, contained more than two samples. The cluster 19 contained five sequences from MSM, three of which reported the use of illicit drugs and two were positive for syphilis (anti-*T. pallidum*). Additionally, cluster 19 also grouped a sequence from São Paulo (Brígido et al., 2011). The two samples characterized as subtype D clustered together (cluster 18).

TABLE 1 | Sociodemographic and behavioral characteristics of 150 cART-naïve subjects according to the HIV-1 most frequent subtypes, Central Brazil.

Variable	N	(%)	Subtype B	Sub-subtype F1	Subtype C
Gender					
Male	93	(62.0)	64 (63.4)	10 (66.7)	7 (58.3)
Female	57	(38.0)	37 (36.6)	5 (33.3)	5 (41.7)
Age (years)					
18–29	49	(32.7)	33 (32.7)	3 (20.0)	6 (50.0)
30–39	51	(34.0)	36 (35.6)	8 (53.3)	2 (16.7)
40 or more	50	(33.3)	32 (31.7)	4 (26.7)	4 (33.3)
Skin color/ethnicity					
White	80	(53.3)	57 (56.4)	8 (53.3)	6 (50.0)
Non-white	70	(46.7)	44 (43.6)	7 (46.7)	6 (50.0)
Educational (years)					
0	4	(2.7)	3 (3.0)	0 (0.0)	0 (0.0)
1–12	117	(78.0)	79 (78.2)	11 (73.3)	9 (75.0)
≥12	29	(19.3)	19 (18.8)	4 (26.7)	3 (25.0)
Monthly income					
<2 minimum wages	20	(13.3)	11 (10.9)	1 (6.7)	2 (16.7)
2–5 minimum wages	97	(64.7)	65 (64.3)	12 (80.0)	6 (50.0)
>5 minimum wages	31	(20.7)	24 (23.7)	2 (13.3)	4 (33.3)
Missing	2	(1.3)	1 (0.1)	0 (0.0)	0 (0.0)
Frequency of alcohol consumption					
None	92	(60.9)	63 (62.4)	8 (53.3)	7 (58.3)
Weekly	51	(35.1)	34 (33.7)	6 (40.0)	3 (25.0)
Daily	7	(4.0)	4 (3.9)	1 (6.7)	2 (16.7)
Illicit drug use					
No	118	(78.7)	80 (79.2)	11 (73.3)	8 (66.7)
Yes, no injecting drugs	28	(18.6)	18 (17.8)	3 (20.0)	4 (33.3)
Yes, injecting drugs	4	(2.7)	3 (3.0)	1 (6.7)	0 (0.0)
Sexual orientation					
Heterosexual, female	57	(38.0)	37 (36.6)	5 (33.3)	5 (41.7)
Heterosexual, male	39	(26.0)	27 (26.8)	3 (20.0)	3 (25.0)
MSM	54	(36.0)	37 (36.6)	7 (46.7)	4 (33.3)
Number of sexual partners in the last 12 months					
0	13	(8.7)	6 (5.9)	2 (13.2)	1 (8.3)
1	77	(51.3)	53 (52.5)	7 (46.7)	6 (50.0)
2–5	40	(26.7)	28 (27.7)	4 (26.7)	4 (33.4)
6–10	4	(2.7)	3 (3.0)	1 (6.7)	0 (0.0)
>10	16	(10.6)	11 (10.9)	1 (6.7)	1 (8.3)
Use of condoms in the last 12 months					
Always	69	(46.0)	51 (50.5)	6 (40.0)	6 (50.0)
Occasionally/Never	81	(54.0)	50 (49.5)	9 (60.0)	6 (50.0)
Presence of TDRM					
Yes	18	(12.0)	14 (13.9)	0 (0.0)	2 (16.7)
No	132	(88.0)	87 (86.1)	15 (100)	10 (83.3)

MSM, men who have sex with men; TDRM, transmitted drug resistance mutations.

DISCUSSION

This phylogenetic study combined detailed clinical and epidemiological data, providing valuable data for surveillance, which allowed the monitoring of HIV-1 variants, TDRM, and associations between sociodemographic characteristics and behavioral sexual groups. It is noteworthy that the study subjects were antiretroviral naïve, and therefore, they were not in virologic suppression at the time of sample collection. This fact, associated

TABLE 2 | Characteristics of the 18 cART-naïve subjects with TDRM.

ID	Age/Gender	Resistance mutations			HIV-1 Subtype	Co-infection
		NRTI	NNRTI	PI		
HET080	50/F	–	V106M	–	BF1	Lifetime syphilis[a]
HET116	23/M	–	K103N	M46I	B	
HSH187	31/M	V75M	–	N88D	B	Lifetime syphilis[a]
HSH430	40/M	L210W, T215D	–	–	B	Hepatitis B[b]
HET446	31/F	–	V106M	–	B	–
HET463	19/M	–	K103N	–	D	–
HSH502	27/M	D67N, K219Q	–	–	C	Lifetime syphilis[a]
HET510	40/F	M41L, T215D	–	M46I, V82T, L90M	B	–
HET521	26/F	K70R	–	–	B	–
HET545	40/M	–	K103N	–	B	Lifetime syphilis[a] Hepatitis B[b]
HET573	37/F	V75M	–	–	B	–
HSH595	28/M	V75M	–	–	B	Hepatitis B[b]
HET607	32/M	V75M	–	–	B	–
HET809	29/M	F77L	–	–	C	–
HET810	31/F	T215S	–	–	B	–
HSH851	21/M	L210W	–	–	B	–
HSH876	27/M	–	K103N	–	B	Lifetime syphilis[a] Hepatitis B[b]
HSH878	22/M	M184V	K103N, P225H	–	B	Lifetime syphilis[a]

HET, sample obtained from heterosexual individual; HSH, sample obtained from men who have sex with men; ID, sample identification; NNRTI, Non-Nucleoside Reverse Transcriptase Inhibitor; NRTI, Nucleoside Reverse Transcriptase Inhibitor; PI, Protease inhibitor. [a]Lifetime syphilis: anti-Treponema pallidum positivity in ELISA. [b]Hepatitis B: anti-HBc total and/or HBsAg seroposivity in ELISA.

with unprotected sexual practices, a multiplicity of sexual partners and a history of sexually transmitted infections (STIs), may be crucial for the maintenance of high HIV transmission rates.

In this study, HIV-1 B subtype was identified in 67.3% of the isolates, followed by recombinant forms, subtypes F1, C, and D. This distribution reflects that found in most Brazilian regions (da Silveira et al., 2012; de Moraes Soares et al., 2014). The frequency of 13.3% (95% CI: 7.9 to 18.8%) of recombinant forms found in this study was similar to that found in previous studies conducted in Central Brazil (16.3% and 14.5%) (Stefani et al., 2007; da Silveira et al., 2012). The absence of the Caribbean non-pandemic subtype B (B_{CAR}) differs from the previous study by Divino et al. (2016), where a frequency of 5.5% from B_{CAR} were detected in Mato Grosso do Sul. Previous studies conducted in a southern region of Brazil identified differences in the distribution of subtypes according to sex and exposure category (De Sa Filho et al., 2005; Dias et al., 2009). The present study is the first conducted in MS addressing this issue, and the lack of association herein can be justified by the high frequency of bisexual behavior (33.9%) reported by homosexual individuals from our cohort, suggesting that the differential transmission of subtypes according to the exposure category is not restricted to the MSM.

In the present study, an intermediate prevalence (12.0%) of TDRM was found, according to the WHO classification (Bennett et al., 2009), which is higher than that found in Northern Brazil (1.0%) (dos Anjos Silva et al., 2016) and is consistent with those found in previous Brazilian studies using similar sequencing technologies (6.8% to 17.2%) (Brindeiro et al., 2003; De Sa Filho et al., 2005; Cardoso et al., 2009; Sprinz et al., 2009; Alencar et al., 2013; Pessôa et al., 2015; Arruda et al., 2018). Recently, among crack cocaine users in Central Brazil, a high prevalence of TDRM was found (58.3%). It is worth noting that only 12 HIV-positive individuals were investigated (Da Silva França et al., 2018).

Recently, one study using massive parallel sequences of Brazilian blood donors found an overall prevalence of TDRM in PR and RT regions of the HIV-1 *pol* gene of 44.5% (Pessôa and Sanabani, 2017). Insufficient data to evaluate the time of HIV-1 infection and conventional sequencing usage may have caused an underestimation of TDRM prevalence (Palmer et al., 2005; Jain et al., 2011; Mohamed et al., 2014). Besides, it has been reported that significant inequalities in access to treatment persists in Brazil, resulting in different impacts on mortality in some groups, such as non-white individuals, or those with poor formal education (Lima et al., 2018).

It is remarkable that 4.0% of virus isolates obtained in this study had multiple mutations that may further influence the response to treatment. K103N, the most frequent resistance mutation observed, is commonly related to decreased susceptibility to efavirenz and nevirapine and the V75M mutation was associated with lamivudine and/or stavudine use (NNRTI). Some studies point out that genotyping tests before initiation of cART for all patients could be cost-effective in Brazil (Sanabani et al., 2011; Luz et al., 2015). However, these tests are still available only to specific populations, such as serodiscordant partners and HIV infected pregnant women.

Although HIV prevalence among MSM increased beyond expectations in Brazil, no difference between TDRM prevalence

TABLE 3 | Cluster confirmation of cART-naïve HIV-1 sequences according to aLRT and genetic distance.

Figure 2			Figure 3				Transmission cluster confirmation
Sequence	**Cluster number**	**aLRT**	**Sequence**	**Cluster number**	**aLRT**	**Genetic distance**	
Subtype HIV-1 B							
HSH645	1	0.98	HSH645	1a	0.99	1.8	Confirmed
HSH684			HSH684				
HSH876			BRMS57				
HSH878			HSH878	1b	0.93		Confirmed
HSH543			HSH876			3.5	
HET220			BRMS171				
HET673			HSH543	1c	0.9	6.0	Not confirmed
HET075			HET220				
			HET673		not confirmed		
			HET075		not confirmed		
HET141	2	0.95	HET141	2	0.94		Confirmed
HET642			HET642			4.2	
HET737			HET737				
HET362	3	0.91	HET362	3	0.93	5.8	Not confirmed
HET622			HET622				
HET219			HET219		not confirmed		
HSH187	4	0.93	HSH187	4	0.95	3.4	Confirmed
HSH595			HSH595				
HET607			BRMS58				
			BRMS14_10				
			HET607				
HSH152	5	1	HSH152	5	1		Confirmed
HSH430			HSH430			3.9	
			BRMS40				
HSH408	6	0.92	HSH408	6a	1	1.3	Confirmed
HSH862			HSH862				
HSH871			HSH871	6b	0.99	1.4	
HET850			BRMS97				Confirmed
			BRMS99				
			HET850		not confirmed		
HET671	7	1	HET671	7	1	2.6	Confirmed
HET438			HET438				
HET517	8	1	HET517	8	1	3.9	
HET797			HET797				Confirmed
HSH703			HSH703				
HET562	9	0.93	HET562	9	0.96	3.6	
HET188			BRMS55				Confirmed
			BRMS05				
			HET188		not confirmed		
HSH143	10	0.93	HSH143		not confirmed		Not confirmed
HET437			HET437		not confirmed		
HSH239	11	0.92	HSH239	11	0.95		Confirmed
HSH066			HSH066			4.4	
HET694			HET694				
HSH881	12	0.96	HSH881	12	0.93	3.3	Confirmed
HET249			HET249				
HET814			HET814		not confirmed		
HSH874			HSH874		not confirmed		
HET367			HET367		not confirmed		

(Continued)

TABLE 3 | Continued

Figure 2			Figure 3				Transmission cluster confirmation
Sequence	**Cluster number**	**aLRT**	**Sequence**	**Cluster number**	**aLRT**	**Genetic distance**	
HET757	13	1	HET757	13	1	1.3	
HSH867			HSH867				
HSH875			HSH875				Confirmed
HSH716			HSH716				
HSH717			HSH717				
HSH53	14	0.98	HSH53	14	0.99	1.9	Confirmed
HSH94			HSH94				
HSH240			HSH240		not confirmed		
HET111	15	0.98	HET111	15	0.96	3.6	
HSH109			HSH109				Confirmed
			MS34				
HET38	16	0.98	HET38	16	0.94	3.7	Confirmed
HET510			HET510				
HET462	17	1	HET462	17	0.98	3.4	Confirmed
HET631			HET631				
HET801			HET801				
			MS02				
			MS46				

Figure 2			Figure 4				Transmission cluster confirmation
Sequence	**Cluster number**	**aLRT**	**Sequence**	**Cluster number**	**aLRT**	**Genetic distance**	
Subtype D							
HET463	18	1	HET463	18	1	3.4	Confirmed
HSH761			HSH761				
Sub-subtype F1							
HSH006	19	1	HSH006	19	0.97	4.0	Confirmed
HSH264			HSH264				
HSH265			HSH265				
HSH387			HSH387				
HSH534			HSH534				
			BR07SP153				
HET256	20	0.99	HET256	20	1	3.6	Confirmed
HET318			HET318				
HET361			HET361				
			BRMS38				
HET140	21	0.92	HET140	21	1	4.8	Not confirmed
HSH739			HSH739				
HET796			HET796				
			BRGO4074				
			BRGO6051				
Recombinant BD							
HET45	22	1					
HET46							
Recombinant BF							
HET643	23	1					
HET122							
HET483							
HET793	24	0.99					
HET353							

HET: sample obtained from heterosexual individual; HSH: sample obtained from men who have sex with men.

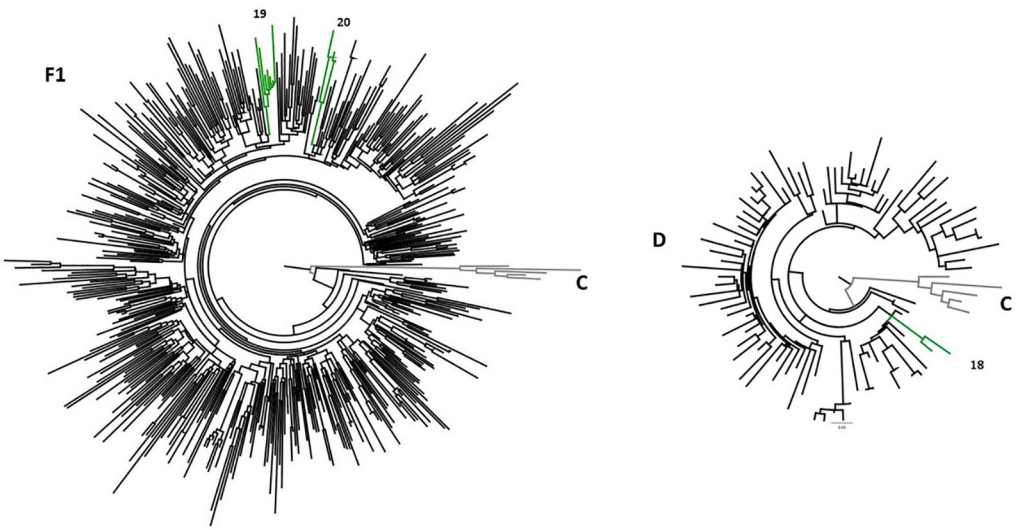


FIGURE 4 | ML phylogenetic tree showing the transmission clusters among HIV-1 subtypes of D and F1 sequences. The confirmed HIV-1 transmission clusters were highlighted in green and numbered according to the previous grouping from **Figure 1**. All clusters present aLRT $\geq$ 0.90 and low mean pairwise genetic distances ($\leq$4.5). The sub-subtype F1 analysis involved 493 PR/RT sequences (15 sequences from the present study, 467 Brazilian reference sequences, 6 non-Brazilian reference sequences, and 5 HIV-1 Subtype C sequences as outgroup). From subtype D analysis, 94 sequences were used as follows: 2 detected in the present study, 17 HIV-1 Brazilian reference sequences, 72 non-Brazilian reference sequences and HIV-1 Subtype C sequences as outgroup).

in homosexuals and heterosexuals was observed in this study. This result may reflect trends of feminization and the increase in heterosexual transmissions observed in Brazil (Brasil, 2017). In contrast, (Bermúdez-Aza et al., 2011) found higher TDRM prevalence in MSM (21.4%) recruited in Brazil by respondent-driven sampling, a particular sampling technique for hard-to-reach populations. As a result, transmission networks of resistance variants may have been selected among these MSM, thus reflecting this prevalence. Due to the higher prevalence of HIV infection in MSM (Kerr et al., 2018) and transgender women in Brazil (Grinsztejn et al., 2017), pre-exposure prophylaxis is recommended by the Brazilian Ministry of Health, who have made efforts to implement it and suggest it may be cost-effective (Luz et al., 2018).

Transmission clusters are frequently defined by low genetic distance (1.0%-4.5%) within cluster sequences and high support phylogenetic clusters (Lewis et al., 2008; Bezemer et al., 2010), herein employing both resources we were able to determine nineteen transmission clusters. However, more recently, transmission network approaches have also been used to this purpose, such as HIV clustering (Wertheim et al., 2014), Cluster picker and Cluster Matcher (Ragonnet-Cronin et al., 2013).

Seventeen transmission clusters were confirmed among subtype B isolates, some of them grouped patients with co-infections. Further evidence suggests that unprotected sexual intercourse and the presence of STIs that cause ulcerative lesions such as syphilis play important roles as cofactors in HIV transmission (Lynn and Lightman, 2004; Karp et al., 2009). This emphasizes the importance of prevention and treatment interventions.

Preventive actions regarding HIV-1 transmission are needed to disrupt the network and to reduce the spread of TDRM, since 29.4% of the clusters (5/17) contained samples with TDRM. Two of these groups were sharing the same substitution, showing the possibility of transmission of resistance between these individuals. Therefore, since 2013 the Brazilian Health Ministry recommendation, following the WHO recommendation, established that all HIV infected individuals should start the treatment to accomplish viral suppression, this being an effective way to reduce the HIV transmission (Brasil, 2013).

Clusters containing sequences from individuals with different sexual behaviors, including homosexual and bisexual contacts, were found in HIV-1 B (clusters 4, 8,11,12, 13, and 15) and D subtypes (cluster 18). Thus, factors such as being a sex worker, having multiple sexual partners, inconsistent condom use, and bisexual behavior may increase exposure to resistant HIV-1 isolates, both in heterosexual and homosexual networks.

The detection of clusters containing Brazilian samples from previous studies in Central-Western and Southeastern Brazil (Brígido et al., 2011; Cardoso et al., 2011; da Silveira et al., 2012) can be explained by the high mobility of the population, reinforcing the possibility of the spreading of infection despite great geographic distances, thus influencing local dynamics of diseases. Therefore, transmission networks and potential links with the different exposure categories should be further investigated in Brazil.

The study has some limitations. We interviewed all individuals face-to-face; consequently, risk behaviors may have been under-reported, leading to potential underestimation of associations

with these variables and TDRM prevalence. Moreover, due to the study design, sample composition may not be representative of Campo Grande-MS epidemic and the absence of time of HIV-1 infection or diagnosis can portray an older epidemic. Even using a very limited number (1.4% from the total number of AIDS cases in Mato Grosso do Sul) of HIV-1 sequences from Mato Grosso do Sul, we were able to detect transmission clusters. However, we could not obtain detailed epidemiological information about the sequences from other Brazilian studies that were in some clusters. On the other hand, these findings enhance the understanding of the HIV-1 genetic characteristics, transmitted drug resistance, and transmission networks, as the research comprises not only individuals with epidemiological features in common but also the spread of strains between homosexuals and heterosexuals.

We highlight the urgent need for increased transmission monitoring of antiretroviral-resistant isolates, aiming for the selection of more effective therapeutic regimens, viral suppression, and hence the interruption of HIV-1 transmission networks. Improved understandings of risks, including potential linkages between sexual exposures among MSM, may contribute to designing preventive interventions and for improving HIV surveillance regarding TDRM in the largest country in Latin America.

AUTHOR CONTRIBUTIONS

MLG and AM-C conceived the presented idea. TT, TFL, MLG, and AM-C discussed the results and wrote the manuscript. TT, SF, GC, and GR collected blood samples and also performed DNA extraction. AL provided medical support. TT, TFL, and MLG performed the experiments. TT, TFL, MLG, and AM-C analyzed the data. All the authors contributed to the final version of the manuscript.

FUNDING

The authors acknowledge CNPq, Fundect-MS 0020/10 (Number 23/200.283/2009), and IOC for providing some financial support. MLG and AM-C are recipient of a CNPq fellowship. TT and TFL are funded by a CAPES Ph.D. fellowship.

ACKNOWLEDGMENTS

We thank Priscila Brunini Zanini, GR and SF who started studying this population in Campo Grande, MS, and recruited a large number of subjects for this research. TT and AM-C also thank MLG for having welcomed us in the Fiocruz laboratory.

REFERENCES

Alencar, C. S., Sabino, E. C., Carvalho, S. M. F., Leao, S. C., Carneiro-Proietti, A. B., Capuani, L., et al. (2013). HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: role of infected blood donors as sentinel populations for molecular surveillance of HIV. *J. Acquir. Immune Defic. Syndr.* 63, 387–392. doi: 10.1097/QAI.0b013e31828ff979

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Arruda, M. B., Boullosa, L. T., Cardoso, C. C., da Costa, C. M., Alves, C. R., de Lima, S. T., et al. (2018). Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): a survey of treatment-naive individuals. *J. Int. AIDS Soc.* 21:e25032. doi: 10.1002/jia2.25032

Bennett, D. E., Camacho, R. J., Otelea, D., Kuritzkes, D. R., Fleury, H., Kiuchi, M., et al. (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 4:e4724. doi: 10.1371/journal.pone.0004724

Bermúdez-Aza, E. H., Kerr, L. R. F. S., Kendall, C., Pinho, A. A., de Mello, M. B., Mota, R. S., et al. (2011). Antiretroviral drug resistance in a respondent-driven sample of HIV-infected men who have sex with men in Brazil. *J. Acquir. Immune Defic. Syndr.* 57(Suppl. 3), S186–S192. doi: 10.1097/QAI.0b013e31821e9c36

Bezemer, D., Van Sighem, A., Lukashov, V. V., Van Der Hoek, L., Back, N., Schuurman, R., et al. (2010). Transmission networks of hiv-1 among men having sex with men in the Netherlands. *AIDS* 24, 271–282. doi: 10.1097/QAD.0b013e328333ddee

Booth, C. L., and Geretti, A. M. (2007). Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. *J. Antimicrob. Chemother.* 59, 1047–1056. doi: 10.1093/jac/dkm082

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de DST, Aids e Hepatites Virais (2013). *Protocolo clínico e Diretrizes Terapêuticas para Manejo da Infecção Pelo HIV em Adultos*.

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de DST, Aids e Hepatites Virais (2017). *Boletim Epidemiologico HIV/Aids*.

Brígido, L. F. M., Ferreira, J. L. P., Almeida, V. C., Rocha, S. Q., Ragazzo, T. G., Estevam, D. L., et al. (2011). Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. *AIDS Res. Hum. Retrovir.* 27, 339–344. doi: 10.1089/aid.2010.0157

Brindeiro, R. M., Diaz, R. S., Sabino, E. C., Morgado, M. G., Pires, I. L., Brigido, L., et al. (2003). Brazilian network for HIV drug resistance surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS* 17, 1063–1069. doi: 10.1097/01.aids.0000060345.12269.d7

Cardoso, L. P., Queiroz, B. B., and Stefani, M. M. (2011). Molecular characteristics of HIV type 1 infection among prisoners from central western Brazil. *AIDS Res. Hum. Retrovir.* 27, 1349–1353. doi: 10.1089/aid.2011.0153

Cardoso, L. P. V., Queiroz, B. B., and Stefani, M. M. (2009). HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naïve patients from Central West Brazil. *J. Clin. Virol.* 46, 134–139. doi: 10.1016/j.jcv.2009.07.009

Da Silva França, D. D., Del-Rios, N. H. A., Dos Santos Carneiro, M. A., Guimarães, R. A., Caetano, K. A. A., Da Guarda Reis, M. N., et al. (2018). HIV-1 infection among crack cocaine users in a region far from the epicenter of the HIV epidemic in Brazil: prevalence and molecular characteristics. *PLoS One* 13:e0199606. doi: 10.1371/journal.pone.0199606

da Silveira, A. A., Cardoso, L. P. V., Francisco, R. B. L., and de Araújo Stefani, M. M. (2012). HIV type 1 molecular epidemiology in pol and gp41 genes among naive patients from Mato Grosso do Sul State, Central Western Brazil. *AIDS Res. Hum. Retrovir.* 28, 304–307. doi: 10.1089/aid.2011.0128

de Medeiros, R. M., Junqueira, D. M., Matte, M. C. C., Barcellos, N. T., Chies, J. A. B., and Almeida, S. E. D. M. (2011). Co-circulation HIV-1 subtypes B, C and CRF31_BC in a drug-naïve population from southernmost Brazil: analysis of primary resistence mutations. *J. Med. Virol.* 83, 1682–1688. doi: 10.1002/jmv.22188

de Moraes Soares, C. M. P., Vergara, T. R. C., Brites, C., Brito, J. D. U., Grinberg, G., Caseiro, M. M., et al. (2014). Prevalence of transmitted HIV-1 antiretroviral resistance among patients initiating antiretroviral therapy in

Brazil: a surveillance study using dried blood spots. *J. Int. AIDS Soc.* 17:19042. doi: 10.7448/IAS.17.1.19042

De Sa Filho, D. J., Sanabani, S., Diaz, R. S., Munerato, P., Brunstein, A., Fusuma, E., et al. (2005). Analysis of full-length human immunodeficiency virus type 1 genome reveals a variable spectrum of subtypes B and F recombinants in São Paulo, Brazil. *AIDS Res. Hum. Retrovir.* 21, 145–151. doi: 10.1089/aid.2005.21.145

Delatorre, E., Silva-de-Jesus, C., Couto-Fernandez, J. C., Pilotto, J. H., and Morgado, M. G. (2017). High HIV-1 diversity and prevalence of transmitted drug resistance among antiretroviral-naive HIV-infected pregnant women from Rio de Janeiro, Brazil. *AIDS Res. Hum. Retrovir.* 33, 68–73. doi: 10.1089/aid.2016.0159

Dias, C. F., Nunes, C. C., Freitas, I. O., Lamego, I. S., De Oliveira, I. M. R., Gilli, S., et al. (2009). High prevalence and association of HIV-1 non-B subtype with specific sexual transmission risk among antiretroviral naïve patients in Porto Alegre, RS, Brazil. *Rev. Inst. Med. Trop. Sao Paulo* 51, 191–196. doi: 10.1590/S0036-46652009000400003

Divino, F., De Corado, A. L. G., Naveca, F. G., Stefani, M. M. A., and Bello, G. (2016). High prevalence and onward transmission of non-pandemic HIV-1 subtype B clades in northern and northeastern Brazilian regions. *PLoS One* 11:e0162112. doi: 10.1371/journal.pone.0162112

dos Anjos Silva, L., Divino, F., da Silva Rêgo, M. O., Lima Lopes, I. G., Nóbrega Costa, C. M., da Silva Pereira, F. C., et al. (2016). HIV-1 genetic diversity and transmitted drug resistance in antiretroviral treatment-naive individuals from Amapá State, Northern Brazil. *AIDS Res. Hum. Retrovir.* 32, 373–376. doi: 10.1089/aid.2015.0280

Fernandes, F. R. P., Zanini, P. B., Rezende, G. R., Castro, L. S., Bandeira, L. M., Puga, M. A., et al. (2015). Syphilis infection, sexual practices and bisexual behaviour among men who have sex with men and transgender women: a cross-sectional study. *Sex. Transm. Infect.* 91, 142–149. doi: 10.1136/sextrans-2014-051589

Freitas, S. Z., Soares, C. C., Tanaka, T. S. O., Lindenberg, A. S. C., Teles, S. A., Torres, M. S., et al. (2014). Prevalence, risk factors and genotypes of hepatitis B infection among HIV-infected patients in the State of MS, Central Brazil. *Braz. J. Infect. Dis.* 18, 473–480. doi: 10.1016/j.bjid.2014.01.005

Gifford, R. J., Liu, T. F., Rhee, S. Y., Kiuchi, M., Hue, S., Pillay, D., et al. (2009). The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* 25, 1197–1198. doi: 10.1093/bioinformatics/btp134

Gräf, T., Passaes, C. P. B., Ferreira, L. G. E., Grisard, E. C., Morgado, M. G., Bello, G., et al. (2011). HIV-1 genetic diversity and drug resistance among treatment naïve patients from Southern Brazil: an association of HIV-1 subtypes with exposure categories. *J. Clin. Virol.* 51, 186–191. doi: 10.1016/j.jcv.2011.04.011

Gräf, T., and Pinto, A. R. (2013). The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 435, 170–178. doi: 10.1016/j.virol.2012.08.048

Grinsztejn, B., Jalil, E. M., Monteiro, L., Velasque, L., Moreira, R. I., Garcia, A. C. F., et al. (2017). Unveiling of HIV dynamics among transgender women: a respondent-driven sampling study in Rio de Janeiro, Brazil. *Lancet HIV* 4, e169–e176. doi: 10.1016/S2352-3018(17)30015-2

Guimarães, M. L., Marques, B. C. L., Bertoni, N., Teixeira, S. L. M., Morgado, M. G., Bastos, F. I., et al. (2015). Assessing the HIV-1 epidemic in Brazilian drug users: a molecular epidemiology approach. *PLoS One* 10:e0141372. doi: 10.1371/journal.pone.0141372

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and Methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 2.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Jain, V., Sucupira, M. C., Bacchetti, P., Hartogensis, W., Diaz, R. S., Kallas, E. G., et al. (2011). Differential persistence of transmitted HIV-1 drug resistance mutation classes. *J. Infect. Dis.* 203, 1174–1181. doi: 10.1093/infdis/jiq167

Karp, G., Schlaeffer, F., Jotkowitz, A., and Riesenberg, K. (2009). Syphilis and HIV co-infection. *Eur. J. Intern. Med.* 20, 9–13. doi: 10.1016/j.ejim.2008.04.002

Kerr, L., Kendall, C., Guimarães, M. D. C., Mota, R. S., Veras, M. A., Dourado, I., et al. (2018). HIV prevalence among men who have sex with men in Brazil: results of the 2nd national survey using respondent-driven sampling. *Medicine* 97(1S Suppl. 1), S9–S15. doi: 10.1097/MD.0000000000010573

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Leigh Brown, A. J. (2008). Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 5:e50. doi: 10.1371/journal.pmed.0050050

Lima, T. A., Beyrer, C., Golub, J. E., Mota, J. C. D., Malta, M. S., Silva, C. M. F. P. D., et al. (2018). Inequalities in HAART uptake and differential survival according to exposure category in Rio de Janeiro, Brazil. *Cad. Saude Publica.* 34:e00009617. doi: 10.1590/0102-311x00009617

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Luz, P. M., Morris, B. L., Grinsztejn, B., Freedberg, K. A., Veloso, V. G., Walensky, R. P., et al. (2015). Cost-effectiveness of genotype testing for primary resistance in Brazil. *J. Acquir. Immune Defic. Syndr.* 68, 152–161. doi: 10.1097/QAI.0000000000000426

Luz, P. M., Osher, B., Grinsztejn, B., Maclean, R. L., Losina, E., Stern, M. E., et al. (2018). The cost-effectiveness of HIV pre-exposure prophylaxis in men who have sex with men and transgender women at high risk of HIV infection in Brazil. *J. Int. AIDS Soc.* 21:e25096. doi: 10.1002/jia2.25096

Lynn, W. A., and Lightman, S. (2004). Syphilis and HIV: a dangerous combination. *Lancet Infect. Dis.* 4, 456–466. doi: 10.1016/S1473-3099(04)01061-8

Machado, L. F. A., Ishak, M. O. G., Vallinoto, A. C. R., Lemos, J. A. R., Azevedo, V. N., Moreira, M. R. C., et al. (2009). Molecular epidemiology of HIV type 1 in Northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. *AIDS Res. Hum. Retrovir.* 25, 961–966. doi: 10.1089/aid.2009.0027

Mohamed, S., Penaranda, G., Gonzalez, D., Camus, C., Khiri, H., Boulmé, R., et al. (2014). Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure. *AIDS* 28, 1315–1324. doi: 10.1097/QAD.0000000000000267

Palmer, S., Kearney, M., Maldarelli, F., Halvas, E. K., Bixby, C. J., Bazmi, H., et al. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 43, 406–413. doi: 10.1128/JCM.43.1.406-413.2005

Pedroso, C., Queiroz, A. T. L., Alcĝntara, L. C., Drexler, J. F., Diaz, R. S., Weyll, N., et al. (2007). High prevalence of primary antiretroviral resistance among HIV-1-infected adults and children in Bahia, a northeast state of Brazil [3]. *J. Acquir. Immune Defic. Syndr.* 45, 251–253. doi: 10.1097/QAI.0b013e318050d8b0

Pessôa, R., and Sanabani, S. S. (2017). High prevalence of HIV-1 transmitted drug-resistance mutations from proviral DNA massively parallel sequencing data of therapy-naïve chronically infected Brazilian blood donors. *PLoS One* 12:e0185559. doi: 10.1371/journal.pone.0185559

Pessôa, R., Watanabe, J. T., Calabria, P., Alencar, C. S., Loureiro, P., Lopes, M. E., et al. (2015). Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus Type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. *Transfusion* 55, 980–990. doi: 10.1111/trf.12936

Ragonnet-Cronin, M., Hodcroft, E., Hué, S., Fearnhill, E., Delpech, V., Brown, A. J. L., et al. (2013). Automated analysis of phylogenetic clusters. *BMC Bioinformatics* 14:317. doi: 10.1186/1471-2105-14-317

Sanabani, S. S., de Pastena, É. R. S., da Costa, A. C., Martinez, V. P., Kleine-Neto, W., de Oliveira, A. C. S., et al. (2011). Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in São Paulo, Brazil. *PLoS One* 6:e25869. doi: 10.1371/journal.pone.0025869

Silva, M. M. G., Telles, F. Q., da Cunha, C. A., and Rhame, F. S. (2010). HIV subtype, epidemiological and mutational correlations in patients from Paraná, Brazil. *Braz. J. Infect. Dis.* 14, 495–501. doi: 10.1590/S1413-86702010000500012

Souza Junior, P. R. B., Borges De Souza Junior, P. R., Szwarcwald, C. L., and Ayres De Castilho, E. (2011). Self-rated health by HIV-infected individuals undergoing antiretroviral therapy in Brazil. *Cad. Saude Publica* 2011(27 Suppl. 1), S56–S66.

Sprinz, E., Netto, E. M., Patelli, M., Lima, J. S., Furtado, J. J. D., da Eira, M., et al. (2009). Primary antiretroviral drug resistance among HIV type 1-infected individuals in Brazil. *AIDS Res. Hum. Retrovir.* 25, 861–867. doi: 10.1089/aid.2009.0012

Stefani, M. M. A., Pereira, G. A. S., Lins, J. A. B., Alcantara, K. C., Silveira, A. A., Viegas, A. A., et al. (2007). Molecular screening shows extensive HIV-1 genetic

diversity in Central West Brazil. *J. Clin. Virol.* 39, 205–209. doi: 10.1016/j.jcv.2007.04.012

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

UNAIDS (2018). *AIDS info [Internet]. The Joint United Nations Programme on HIV/AIDS*. Available at: http://aidsinfo.unaids.org/#

Wertheim, J. O., Leigh Brown, A. J., Hepler, N. L., Mehta, S. R., Richman, D. D., Smith, D. M., et al. (2014). The global transmission network of HIV-1. *J. Infect. Dis.* 209, 304–313. doi: 10.1093/infdis/jit524

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 February 2019
doi: 10.3389/fmicb.2019.00145

Reduction of HIV-1 Reservoir Size and Diversity After 1 Year of cART Among Brazilian Individuals Starting Treatment During Early Stages of Acute Infection

Thaysse Ferreira Leite[1†], Edson Delatorre[1†‡], Fernanda Heloise Côrtes[1], Ana Cristina Garcia Ferreira[2], Sandra Wagner Cardoso[2], Beatriz Grinsztejn[2], Michelle Morata de Andrade[2], Valdilea Gonçalves Veloso[2], Mariza Gonçalves Morgado[1] and Monick Lindenmeyer Guimarães[1]*

[1] Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [2] Laboratório de Pesquisa Clínica em DST e AIDS, Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Kazutaka Terahara,
National Institute of Infectious Diseases (NIID), Japan
Guido van Marle,
University of Calgary, Canada

***Correspondence:**
Edson Delatorre
delatorre.ioc@gmail.com;
edsonod@ioc.fiocruz.br

[†]Joint first authors

[‡]**Present address:**
Edson Delatorre,
Laboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 23 August 2018
Accepted: 21 January 2019
Published: 11 February 2019

Citation:
Leite TF, Delatorre E, Côrtes FH, Ferreira ACG, Cardoso SW, Grinsztejn B, de Andrade MM, Veloso VG, Morgado MG and Guimarães ML (2019) Reduction of HIV-1 Reservoir Size and Diversity After 1 Year of cART Among Brazilian Individuals Starting Treatment During Early Stages of Acute Infection. Front. Microbiol. 10:145. doi: 10.3389/fmicb.2019.00145

The aim of early combined antiretroviral therapy (cART) of HIV is to limit the seeding of the viral reservoir during the initial phase of infection and, consequently, decrease intrahost viral diversity. Here, we assessed the effect of early cART on size and complexity of the proviral reservoir. Peripheral blood mononuclear cell (PBMC) and plasma samples were obtained from ten HIV-infected Brazilian individuals diagnosed at the acute phase of infection, before (PRE_{ART}) and 12 months ($M12_{ART}$) after suppressive cART. HIV proviral reservoir size was determined by quantitative real-time PCR; intrahost viral diversity of the *env* C2-V3 region was assessed by single genome amplification or next-generation sequencing in PBMC and plasma, respectively. Mean nucleotide diversity (π) and normalized Shannon entropy (H_{SN}) were used to infer the complexity of the viral population. Compared to PRE_{ART}, $M12_{ART}$ saw an immunological recovery with a gain of ∼200 $CD4^+$ T cells (P = 0.008) and a normalization of the CD4/CD8 ratio [1.0 (IQR: 0.88–1.18), P = 0.016], as well as a significant decrease in HIV-1 RNA (∼4 log, P = 0.004) and DNA (∼1 log, P = 0.002) levels. The median time to achieve viral suppression was 3 months (IQR: 2.8–5.8 months). The high intermixing between sequences from both visits suggests that the HIV-1 DNA reservoir remained remarkably stable under cART. After 1 year of cART, there was a minor reduction in proviral π (Pre_{ART} = 0.20 vs. $M12_{ART}$ = 0.10; P = 0.156) but a significant decrease in H_{SN} (Pre_{ART} = 0.41 vs. $M12_{ART}$ = 0.25; P = 0.019). We found no correlation between π or H_{SN} at Pre_{ART} and the rate of HIV DNA decay, T $CD4^+$ counts, or CD4/CD8 ratio at $M12_{ART}$. Based on a small cohort of Brazilian infected individuals under early cART and analyses of the *env* region, 1 year of follow-up suggested

a reservoir size reduction, allowed a significant decrease of HIV-1 complexity, and achieved immunological restoration regardless of the initial HIV-1 plasma viral load, $CD4^+$ T cell counts, or HIV-1 subtype. However, further studies in the Brazilian setting aiming a longer follow-up and larger cohort are required in this field.

Keywords: HIV-1, reservoir, diversity, early cART, acute infection

INTRODUCTION

Combined antiretroviral therapy (cART) suppresses HIV-1 replication and reduces morbidity and mortality, but does not eradicate HIV-1 infection, as a low but persistent level of HIV-1 can still be detected in plasma and cell reservoirs (Chun et al., 1997; Kiselinova et al., 2015; Ghosn et al., 2018). Latent infected resting memory $CD4^+$ T lymphocytes are the best known HIV-1 reservoir, which is established already during early infection. This reservoir includes cells with an integrated copy of the HIV-1 genome that is not expressed while the cells remain in a resting state (Chun et al., 2002; Douek et al., 2002) and is maintained mainly by the cells' clonal expansion (Chomont et al., 2009, 2011; von Stockenstrom et al., 2015). However, it has been suggested that persistent virus replication may be an important contributor to its maintenance (Buzón et al., 2010; Yukl et al., 2010), particularly in lymphoid tissue sanctuary sites (Lorenzo-Redondo et al., 2016). Elimination/reduction of this latent reservoir represents a great hope for curing HIV-1 infection. As the reservoir is established during the acute phase of infection, early cART has been proposed as a means to restrict reservoir size and genetic complexity (Lori et al., 1999; Strain et al., 2005; Chomont et al., 2009; Josefsson et al., 2013).

HIV-1 sequence diversity is limited by the "genetic bottleneck effect" during sexual transmission, which selects viruses with the highest overall fitness (Carlson et al., 2014) but results in a more homogeneous viral population. However, in cohorts of men who have sex with men (MSM), the selection for fitter variants appears less stringent, resulting in infection being established by multiple founding viruses (Gottlieb et al., 2008; Li et al., 2010). Accordingly, the complexity of sequence diversity during early HIV-1 infection may affect the efficacy of cART in decreasing the reservoir, mainly due to immune and therapeutic escape mutations. Another potential escape route is through viral recombinants (Abrahams et al., 2009; Kearney et al., 2009; Bar et al., 2010; Li et al., 2010; Batorsky et al., 2011; Novitsky et al., 2011). Therefore, investigating the diversity of viral variants during primary infection may help evaluate viral evolution and predict clinical outcomes.

Residual replication has been proposed as a mechanism that maintains the HIV-1 reservoir during cART. Failure to block the viral replication cycle, enables renewed cellular infections and continuous replenishment of the HIV-1 DNA reservoir (Josefsson et al., 2013). Individuals who initiated cART during early infection, may present low viral diversity in peripheral blood mononuclear cell (PBMC) and plasma reservoirs after years on suppressive therapy, and the slight sign of viral replication indicates that the reservoir was maintained by homeostatic cell proliferation (Josefsson et al., 2013). A recent study demonstrated a faster decline in HIV DNA levels in early cART-treated patients with homogeneous viral populations after 6 months of therapy (Wang et al., 2017).

Understanding the dynamics of long-lived cellular HIV-1 reservoirs in individuals treated during primary infection, can direct the choice of long-term treatment regimens to achieve post-treatment control of HIV-1 in Brazil. Here, we investigated the effect of early cART during the initial Fiebig stages of HIV-1 acute infection on the size, diversity, and complexity of HIV-1 total DNA in the PBMC reservoir by examining HIV-1 populations prior and 1 year after cART initiation.

MATERIALS AND METHODS

Subject Characteristics

A cohort of Brazilian individuals presenting recent HIV-1 infection has been followed-up since August 2013 at the Instituto Nacional de Infectologia Evandro Chagas (INI-FIOCRUZ). These individuals initiated cART immediately after diagnosis and have been described elsewhere (Ferreira et al., 2017). For the present study, we selected only individuals, who started cART during the acute phase (Fiebig II–V) (Fiebig et al., 2003) of HIV-1 infection ($n = 10$). Participants were recruited between December 2014 and October 2015, and had at least 1 year of successful cART after that. PBMC and plasma samples were obtained at the baseline visit (PRE_{ART}) and 12 months after cART onset ($M12_{ART}$), and were stored until use. The processing of all HIV samples was performed in accordance with institutional standard biosecurity and safety procedures at biosafety level 2. The study was approved by the INI Ethical Review Board (approval number 36859614.8.0000.5262), and all subjects gave written informed consent in accordance with the Declaration of Helsinki.

$CD4^+$ and $CD8^+$ T Cell Counts and HIV-1 RNA Quantification

Peripheral blood $CD4^+$ and $CD8^+$ T cell counts were determined by flow cytometry using the MultiTest TruCount-Kit and MultiSet software on a FACSCalibur flow cytometer (BD Biosciences, United States). HIV-1 RNA in plasma was measured by the Abbot Real-Time HIV-1 Assay, whose lower limit of detection was 40 copies/mL (Abbott Laboratories, Germany).

HIV-1 Total DNA Measurement in PBMCs

Total cellular DNA was extracted from cryopreserved PBMCs (1×10^7 cells) obtained at PRE_{ART} and $M12_{ART}$ using the QIAamp DNA Mini Kit (Qiagen, Germany). Cell-associated HIV-1 DNA was quantified using the Generic HIV® DNA

Cell Kit (Biocentric, France), following the manufacturer's instructions. The assay's lower limit of detection was 40 HIV DNA copies/10^6 cells.

HIV-1 DNA Single Genome Amplification (SGA)

Proviral DNA was extracted from PBMCs using the QIAamp DNA Blood Mini Kit (Qiagen, United States) according to the manufacturer's instructions. HIV-1 quasispecies was obtained by SGA of a 552-bp fragment from the C2-V3 *env* region through nested PCR using Platinum Taq DNA polymerase (Invitrogen, United States) as described elsewhere (Delwart et al., 1993). Considering a Poisson distribution, at a dilution in which approximately 30% of amplicons are positive, a single amplifiable molecule is present about 80% of the time (Palmer et al., 2005). The PCR products were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, United Kingdom). Sequences were obtained using the ABI BigDye Terminator v.3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, United States) on an ABI 3130 Genetic Analyzer (Applied Biosystems). Sequences were assembled and edited using SeqMan 7.0 software (DNASTAR Inc., United States). APOBEC3G/F-mediated hypermutations were revealed by Hypermut software (Rose and Korber, 2000) and sequences showing ambiguous bases were excluded.

HIV-1 *env* RNA Haplotypes Reconstruction From NGS Data

Viral RNA from plasma samples collected at PRE_{ART} (baseline) was extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Germany). The cDNA was obtained by reverse-transcribed PCR using the SuperScript™ III Reverse Transcriptase (Invitrogen, United States) and was then subjected to nested PCR for amplification of the *env* gene as described above. The resulting amplicons were made into a library using the Nextera® XT DNA Library Prep Kit with unique barcodes from the Nextera® XT Index Kit (Illumina, United States), following the manufacturer's instructions. DNA sequencing was performed on a MiSeq instrument using MiSeq Reagent Nano Kit, v2 (500 cycles; Illumina, United States). Demultiplexed reads were trimmed to remove adaptors, low-quality bases (Q <25), and short reads (<100 bp), and then mapped against single-genome amplification consensus sequences from each patient using Geneious software v.9.1.8 (Kearse et al., 2012) with high mapping quality (MAQ ≥30). Alignment regions with at least 500× coverage were used for haplotype reconstruction with QuasiRecomb 1.2 (Marz et al., 2014), employing the flag "-conservative" to increase specificity. Only haplotypes with frequencies ≥1% were used for further analysis.

HIV-1 Subtyping

Sequences were aligned with HIV-1 reference sequences from the Los Alamos database[1] using ClustalW in MEGA 6 and were manually edited. The final *env* alignment covered positions 6,840–7,372 relative to the HXB2 genome. Maximum-likelihood (ML) phylogenetic trees were reconstructed with PhyML 3.0 (Guindon et al., 2010) using the most appropriate nucleotide substitution model selected with jModeltest v. 3.7 (Darriba et al., 2012). The approximate likelihood-ratio test (aLRT) was used to estimate the confidence of branching on the tree.

Analyses of Viral Diversity

Complexity of the intrahost viral population was assessed through two diversity measures. The mean nucleotide diversity (π), an abundance-based functional index representing the average number of nucleotide differences between any two representatives of the population, was calculated in MEGA7 (Kumar et al., 2016). The normalized Shannon entropy (H_{SN}), an abundance-based index that measures viral population diversity based on haplotype frequencies, was calculated in the R package vegan (Oksanen et al., 2017), after sample rarefaction to correct for bias in sample size (Gregori et al., 2016).

Phenotypic Prediction of Co-receptor Usage

The quasispecies viral tropism was predicted based on the V3 amino acid sequence through the Geno2pheno algorithm (available at http://coreceptor.geno2pheno.org). The false positive rate cut-off was 10% for DNA, whereas RNA sequences were classified as CCR5 or non-CCR5-using viruses as described previously (Lengauer et al., 2007; Hayashida et al., 2017).

Statistical Analyses

All statistical analyses were performed in GraphPad Prism v6 (GraphPad Software, United States). Variables between groups (unpaired) or from the same group from different visits (paired) were compared using the Mann–Whitney *U*-test or Wilcoxon test, respectively. Association between variables was evaluated using the Spearman's rank correlation. Linear regression was used to calculate the rate of HIV-1 DNA variation between visits in each subject. *P*-values <0.05 were considered statistically significant.

Sequence Availability

SGA *env* sequences have been submitted to GenBank under accession numbers MH765045–MH765329. NGS data have been deposited to the NCBI BioProject database under accession number PRJNA487221.

RESULTS

Clinical, Epidemiological, and Immunological Characteristics of Early-Treated Individuals

The study included ten HIV-1-infected Brazilian individuals diagnosed during the acute phase of infection, with a median age at diagnosis of 28 years [interquartile range (IQR): 26–42 years]. All participants were MSM and started cART immediately after HIV-1 diagnosis. The cART regimen varied

[1] https://www.hiv.lanl.gov/

among participants, however, all included two co-formulated nucleoside reverse transcriptase inhibitors (lamivudine plus tenofovir) in combination with a non-nucleoside reverse transcriptase inhibitor (efavirenz, n = 5), a protease inhibitor (fosamprenavir/atazanavir, n = 2 each), or an integrase inhibitor (raltegravir, n = 1).

Prior to cART initiation (Pre_{ART} visit), median $CD4^+$ and $CD8^+$ T cell counts of all individuals were 634 and 1473 cells/mm^3 (IQR: 420–886 for $CD4^+$ and 526–1840 for $CD8^+$ cells), respectively, and the CD4/CD8 ratio was 0.52 (IQR: 0.42–0.83). After 12 months on cART ($M12_{ART}$), we observed a significant immunological recovery, with median $CD4^+$ counts of 836 cells/mm^3 (IQR: 719–1122) (**Figures 1A**, **2**). This meant an average increase of ∼200 $CD4^+$ T cells (P = 0.008) and a significant rise in the CD4/CD8 ratio [median of 1.0 (IQR: 0.88–1.18), P = 0.016] (**Figures 1B**, **2**).

The immunological improvement observed 12 months after cART initiation was accompanied by a significant decrease in the median levels of HIV-1 total DNA in PBMCs and HIV-1 RNA in plasma (**Figures 1C,D**). The plasma HIV-1 viral load decreased drastically (∼4 log) between PRE_{ART} and $M12_{ART}$ visits (from 5.86 to <1.6 log copies/mL, P = 0.004), whereas HIV-1 DNA in the PBMC compartment displayed a modest yet significant reduction (∼1 log, from 3.04 to 2.18 log copies/10^6 PBMC, P = 0.002). The median time to achieve viral suppression (<1.6 log HIV RNA copies/mL) after the onset of cART was approximately 3 months (IQR: 2.8–5.8 months), and only one individual (IVA41) exhibited a minor viral rebound (2.66 log HIV RNA copies/mL) within the 12 months of follow-up (**Figure 2**). It should be noted, however, that sustained viral loads were recovered after that (data not shown).

Sequence Analysis

To understand the effect of early cART initiation on HIV-reservoir diversity and complexity, we analyzed the intrahost viral population by SGA of the HIV-1 *env* gene before and after cART. Similar quantities of SGA HIV-1 proviral sequences were obtained from both visits' samples: 15 (IQR: 15–16) in Pre_{ART} and 14.5 (IQR: 11–16) in $M12_{ART}$ (P = 0.14). Additionally, we used NGS to evaluate HIV diversity in the individuals' plasma samples at the Pre_{ART} visit. The median coverage per sample was 8,182 (IQR: 4,755–10,904) reads per base (**Supplementary Figure S1**), and after assembly between one to eight HIV-1 *env* haplotypes were reconstructed per sample.

All sequences obtained from proviral DNA and plasma RNA branched together in highly supported monophyletic clusters (aLRT = 1) by subject (**Figure 3A**); suggesting that in all subjects infection resulted from a single or limited number of closely related viral variants. It is noteworthy that in four subjects (IVA32, IVA21, IVA17, and IVA37), the HIV-1 RNA sequences

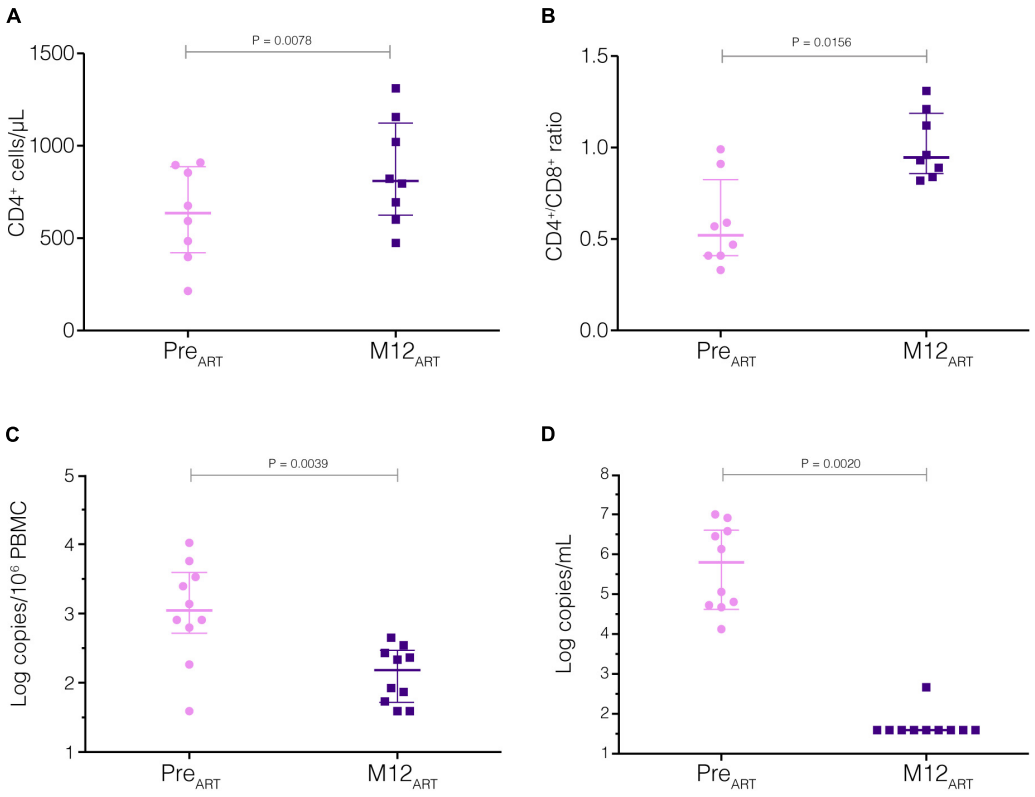


FIGURE 1 | Immunological and virological measurements before and after cART initiation. $CD4^+$ T cell counts **(A)**, $CD4^+/CD8^+$ ratios **(B)**, HIV-1 proviral load in PBMCs **(C)**, and HIV-1 viral load in plasma **(D)** were measured at Pre_{ART} and $M12_{ART}$ visits (pink circles and blue squares, respectively). P-values <0.05 were considered statistically significant.

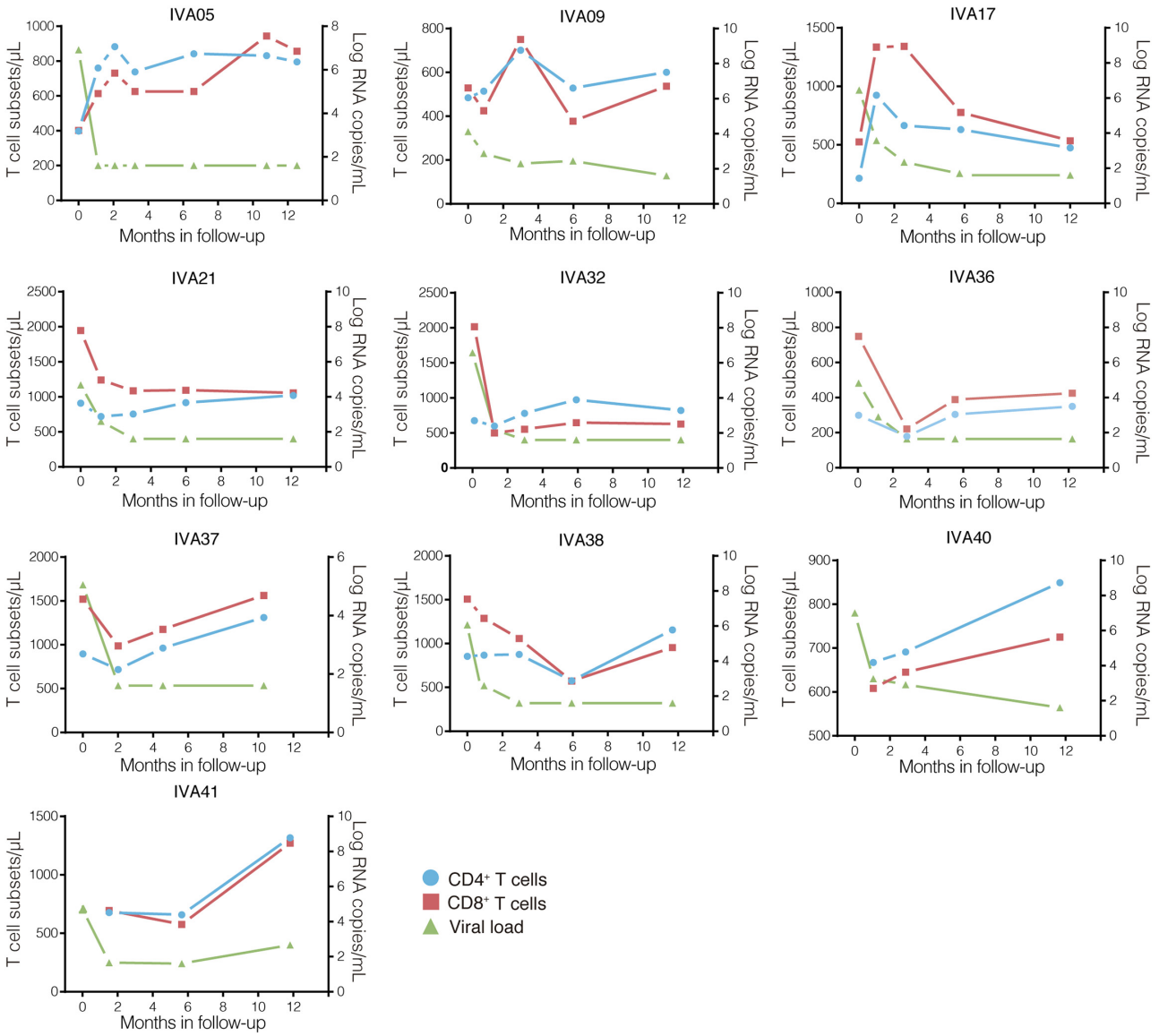


FIGURE 2 | Clinical follow-up of the ten early-treated individuals. T cell counts (cells/μL; $CD4^+$, blue circles; $CD8^+$, red squares) and plasma RNA viral loads (log copies/mL, green triangles) values over time after cART onset (months) are shown on the left and right *Y*-axis, respectively. The individuals' identifications are indicated at the top of each graph.

were the closest to the common ancestor of the subjects' HIV-1-infecting lineage and outside the main cluster comprising the proviral variants. In the remaining subjects, the HIV-1 RNA sequences were intermixed with proviral DNA sequences. HIV-1 subtype B was the most frequent variant detected ($n = 6$), followed by subtypes F1 and C ($n = 2$ in each). R5-tropic viruses dominated the intrahost HIV-1 population in plasma and PBMCs at both visits, with only two individuals (IVA21 and IVA40) presenting a low frequency (6%) of X4-tropic viral clones in PBMCs at Pre_{ART} (**Table 1** and **Figure 3A**). APOBEC3G/F-induced hypermutations were identified at Pre_{ART} in proviral sequences from only three individuals (IVA09, IVA36, and IVA40), with frequencies of 6–12%.

To address the influence of early cART on reservoir complexity, we measured the average pairwise nucleotide differences of the viral population (π) and the uniformity of the haplotype distribution (H_{SN}) of viral DNA and RNA from each visit. Hypermutated sequences were excluded from these analyses. All individuals exhibited low π values (<1.5%) despite the Fiebig stage at cART initiation. We observed a minor decline in median proviral π after 1 year of cART (Pre_{ART} = 0.20, IQR: 0.10–0.58 vs. $M12_{ART}$ = 0.10, IQR: 0.20–0.33; $P = 0.156$) (**Figure 3B**). There was, however, a strong positive correlation between proviral and plasma π values before cART onset ($r = 0.77$; $P = 0.0157$, **Supplementary Figure S2**), which was not observed at $M12_{ART}$ nor when comparing π values from

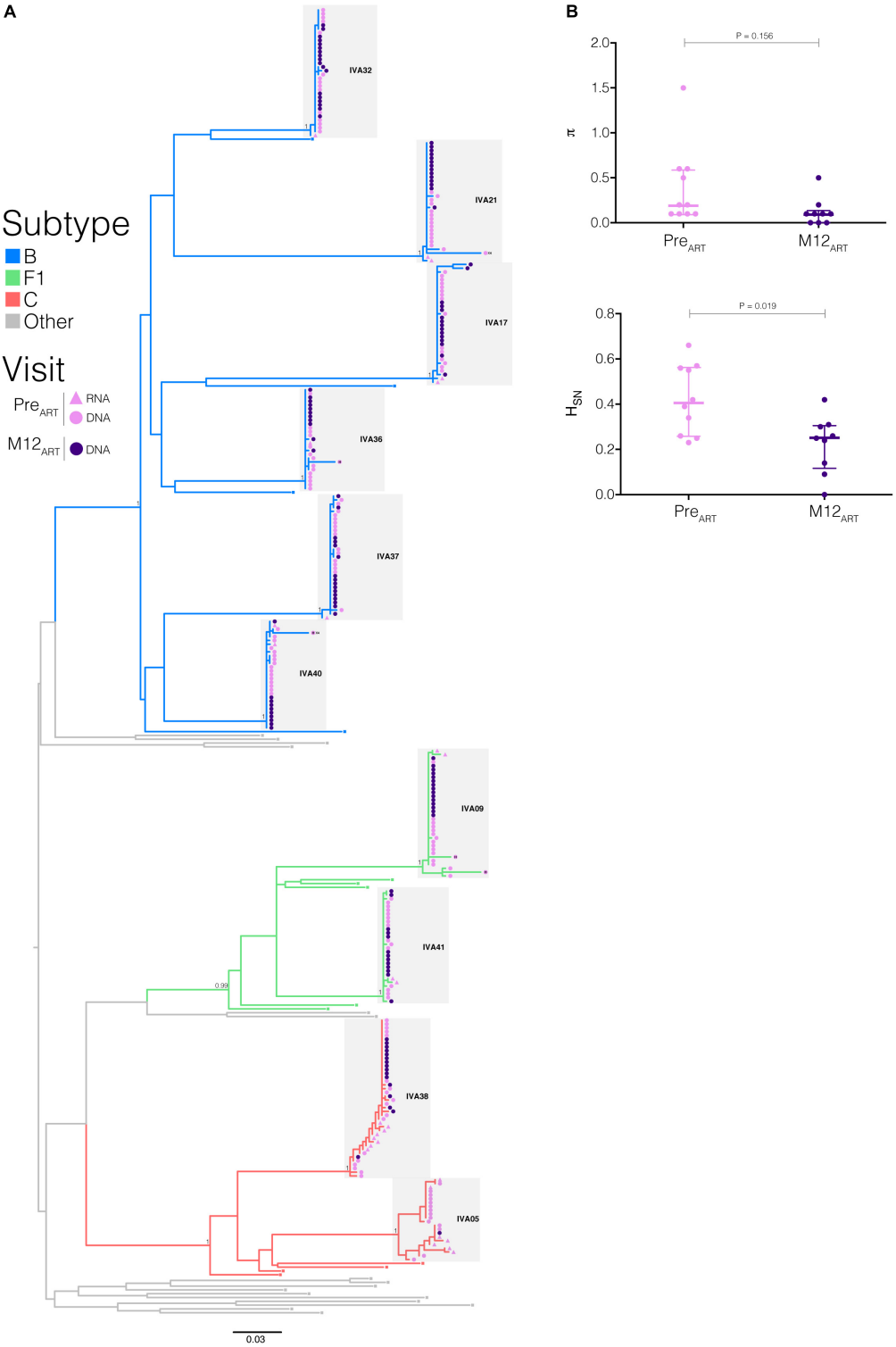


FIGURE 3 | Impact of early cART initiation on intrahost HIV-1 population diversity. **(A)** ML phylogenetic tree of the *env* sequences from Pre_{ART} and $M12_{ART}$ visits. Tips' shapes represent the viral compartment (proviral DNA, circles; plasma RNA, triangles) and are color-coded according to the visit (Pre_{ART}, pink; $M12_{ART}$, purple). The branches' colors agree with the subtype assignment as indicated in the legend. Clusters from each individual are indicated by shaded gray boxes. Branch supports (aLRT-SH) are indicated at key nodes. Tips shapes marked with an "H" indicate the presence of APOBEC3G-mediated G to A hypermutations and X4 labels highlight X4-tropic sequences. Horizontal branch lengths are proportional to the bar at the bottom indicating nucleotide substitutions per site. **(B)** Mean nucleotide diversity (π) and normalized Shannon entropy (H_{SN}) indices were calculated from the proviral *env* sequences obtained at Pre_{ART} and $M12_{ART}$ visits. Thick and thin lines represent the median and interquartile ranges, respectively. *P*-values <0.05 were considered statistically significant.

TABLE 1 | Virological characteristics of HIV-1 early-infected individuals.

Patient ID	Fiebig stage	Subtype	Visit[1]	PMBC						Plasma				
				Log_{10} HIV DNA[2]	*env* clones	π (%)	H_{SN}	Unique sequences (%)	R5 tropic (%)	Log_{10} HIV RNA[3]	*env* haplotypes	π (%)	H_{SN}	R5 tropic (%)
IVA05	II	C	Pre_{ART}	<1.6	14	1.6	0.6	46.6	100	6.9	7	2.7	0.2	100
			$M12_{ART}$	<1.6	1	–	0.0	100	100	<1.6	–	–	–	–
IVA09	V	F1	Pre_{ART}	2.8	17	0.5	0.4	26.6	100	4.1	2	1.0	0.0	100
			$M12_{ART}$	1.7	15	0	0.0	6.7	100	<1.6	–	–	–	–
IVA17	V	B	Pre_{ART}	4.0	15	0.1	0.3	26.6	100	6.5	2	0.3	0.1	100
			$M12_{ART}$	2.5	15	0.5	0.2	26.6	100	<1.6	–	–	–	–
IVA21	V	B	Pre_{ART}	2.9	16	0.6	0.2	25	93.7	4.7	2	0.3	0.1	100
			$M12_{ART}$	1.9	14	0	0.1	14.3	100	<1.6	–	–	–	–
IVA32	V	B	Pre_{ART}	3.4	15	0.1	0.3	20	100	6.6	1	–	0.0	100
			$M12_{ART}$	2.4	18	0.1	0.3	22.2	100	<1.6	–	–	–	–
IVA36	IV	B	Pre_{ART}	2.9	15	0.1	0.4	21.4	100	4.8	1	–	0.0	100
			$M12_{ART}$	1.9	11	0.1	0.3	27.3	100	<1.6	–	–	–	–
IVA37	V	B	Pre_{ART}	3.1	17	0.2	0.6	41.2	100	50	1	–	0.0	100
			$M12_{ART}$	2.7	16	0.1	0.2	25	100	<1.6	–	–	–	–
IVA38	IV	C	Pre_{ART}	3.5	18	0.6	0.7	50	100	6.1	8	0.5	0.4	100
			$M12_{ART}$	2.4	16	0.2	0.4	37.5	100	<1.6	–	–	–	–
IVA40	IV	B	Pre_{ART}	3.8	17	0.2	0.6	37.5	94.1	7	2	0.3	0.1	100
			$M12_{ART}$	2.3	10	0	0.1	20	100	<1.6	–	–	–	–
IVA41	V	F1	Pre_{ART}	2.3	15	0.1	0.2	26.6	100	4.7	2	0.3	0.1	100
			$M12_{ART}$	<1.6	13	0.1	0.3	30.8	100	2.7	–	–	–	–

[1] Pre_{ART}: before cART start, $M12_{ART}$: 12 months on suppressive cART; [2] expressed as HIV-1 DNA copies/10^6 PBMC; [3] expressed as HIV-1 RNA copies/mL.

HIV DNA between visits. The overall median H_{SN} estimated for the HIV-1 proviral population decreased significantly between Pre_{ART} and $M12_{ART}$ visits (0.41, IQR: 0.28–0.56 vs. 0.25, IQR: 0.14–0.30; $P = 0.019$) (**Figure 3B**). However, no correlation was found when comparing the H_{SN} values calculated from different compartments or visits.

In spite of relatively low π values at Pre_{ART}, the values varied among participants (range: 0.1–1.5%). A comparison revealed no significant correlation between π or H_{SN} at Pre_{ART} with immunological and virological parameters, such as rate of HIV DNA decay, T $CD4^+$ cell change, or the CD4/CD8 ratio, at $M12_{ART}$ (**Figure 4**).

DISCUSSION

A latent HIV-1 reservoir in resting memory CD4+ T cells is recognized as the major barrier to HIV-1 eradication, as its establishment and long-term persistence enables renewed viremia after treatment failure or interruption (Richman et al.,

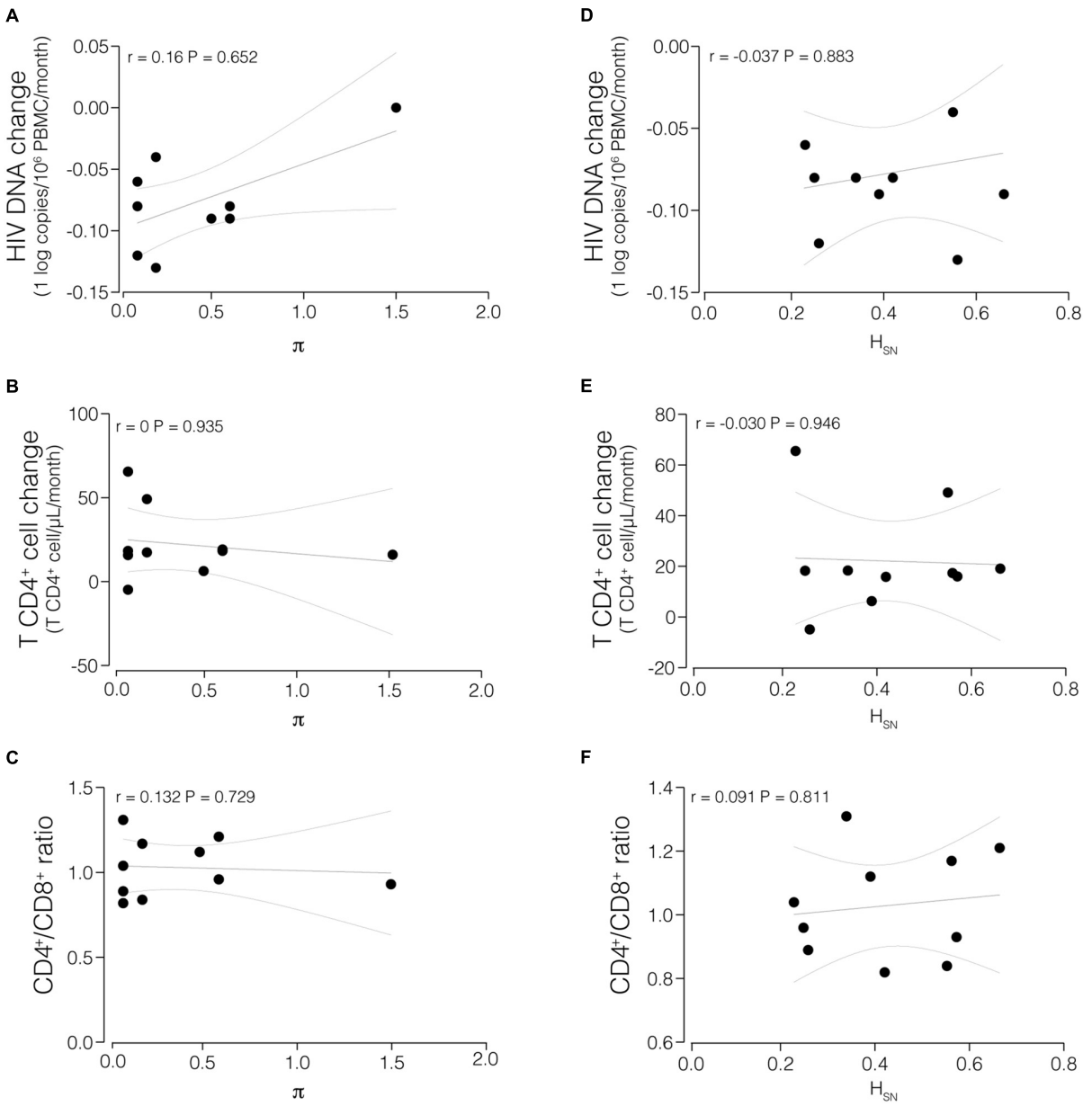


FIGURE 4 | Correlations between proviral HIV-1 diversity indices and immunological and virological measurements. Mean nucleotide diversity (π) of the proviral population at Pre_{ART} was compared with HIV decay in PBMCs **(A)**, T $CD4^+$ cell change **(B)**, and $CD4^+/CD8^+$ ratio in $M12_{ART}$ **(C)**. Normalized Shannon entropy (H_{SN}) of the proviral population at Pre_{ART} was compared with HIV decay in PBMCs **(D)**, T $CD4^+$ cell change **(E)**, and $CD4^+/CD8^+$ ratio in $M12_{ART}$ **(F)**. *P*-values <0.05 were considered statistically significant.

2009). Early cART has been proposed as a means for achieving long-term control of viral replication upon cART interruption, by delaying the viral rebound and potentially inducing a post-treatment controller status (Archin et al., 2012; Avettand-Fènoel et al., 2016). Here, we analyzed ten acutely HIV-infected patients starting cART during the early Fiebig stages, before and 1 year after treatment, to investigate the effect of early therapy on the size and complexity of HIV-1 total DNA in the PBMC reservoir.

Benefits of early cART include a reduction in residual viral replication, a restraint of viral diversity and reservoir development, and accelerated immune restoration (Yerly et al., 2000; Ngo-Giang-Huong et al., 2001; Delwart et al., 2002; Hecht et al., 2006). We observed significant immunological restoration in all subjects 1 year after cART onset during acute HIV infection, with increases of both $CD4^+$ T cell counts and the CD4/CD8 ratio (∼1.0), in agreement with previous results (Hoenigl et al., 2016). The rise of the CD4/CD8 ratio was previously linked with reduced levels of HIV DNA in peripheral blood cells (Chun et al., 2002), mainly if cART was established during primary HIV-1 infection (Hocqueloux et al., 2013). Thus, initiating cART during the acute phase of HIV infection offers an opportunity to reduce HIV reservoirs and achieve optimal immune reconstitution. Nevertheless, a recent study (Colby et al., 2018) has demonstrated that a CD4/CD8 ratio <1 is associated with a rapid viral rebound following treatment interruption, in spite of early cART initiation and sustained undetectable viral load along treatment follow-up. The immunological recovery was accompanied here by a reduction in both HIV RNA and DNA levels. The median viral load decay time to undetectable levels was about 3 months after cART start, similar to previous studies (Pilcher et al., 2004; Hoenigl et al., 2016), indicating that a rapid and efficient virologic suppression was achieved in this cohort. The median HIV-1 DNA load of our subjects at baseline visit was ∼3 log copies/10^6 PBMCs, similar to the one described in the ANRS PRIMO cohort (Schiffer et al., 2004). After 1 year of suppressive cART, HIV-1 DNA levels in PBMCs were reduced to ∼2 log copies/10^6 PBMCs, reaching similar levels as those of HIV-1-infected patients, who had started intensive and standard cART during primary HIV-1 infection (Hocqueloux et al., 2013; Chéret et al., 2015).

A reduction of the HIV-1 latent reservoir may help HIV-1-infected patients reach at least a transient drug-free remission of their disease (Hocqueloux et al., 2013; Sáez-Cirión et al., 2013). Lower levels of HIV DNA in PBMCs at antiretroviral treatment interruption have been associated with a longer time to treatment resumption (Piketty et al., 2010). Despite the reduction in HIV-1 total DNA in the PBMC reservoir observed in our study, the values were still higher (median 2.18 log copies/10^6 PBMCs) than in post-treatment controllers (PTC) from the VISCONTI cohort (median 1.71 log copies/10^6 PBMCs), for whom long-term viral control was maintained after treatment interruption (Sáez-Cirión et al., 2013). The PTC's median time to plasma viral load becoming undetectable was similar to the one calculated in this study; however, median cART duration was much longer (36.5 months in PTC vs. 12 months in this study). Moreover, a study conducted in Thailand showed that total HIV DNA mean values in PBMCs from early-treated acutely HIV-infected individuals decreased over time achieving very low levels after 3 years of cART (Ananworanich et al., 2016). Thus, with longer follow-up of the patients included in this study, total HIV DNA may reduce even further in the PBMC reservoir.

Here, analysis of HIV *env* sequences from plasma and PBMCs revealed very high population intermixing between the clones obtained before and after cART onset, indicating that the HIV-1 DNA reservoir remained remarkably stable. This result agrees with the findings of Josefsson et al. (2013), whereby viral populations in both pre-therapy plasma and cells isolated after suppressive cART initiated during early/acute infection were nearly monomorphic. The absence of genetic changes indicates that maintenance of the HIV reservoir during suppressive cART, as observed in this study, was the result of homeostatic cell proliferation rather than of ongoing viral replication. This is consistent with the results of other studies analyzing HIV reservoir sequences but from longer time intervals than those reported here (Chomont et al., 2009; Kearney et al., 2014; von Stockenstrom et al., 2015; Brodin et al., 2016; Van Zyl et al., 2017). However, the absence of divergence in the PBMC HIV DNA reservoir does not exclude the possibility that ongoing residual replication could be occurring in sanctuary sites, especially in lymphoid tissue (Lorenzo-Redondo et al., 2016). A more robust sampling scheme, assessing distinct compartments and over a longer time frame should be performed to increase the chances of identifying putatively evolving viruses.

Recent studies have demonstrated that patients, who were treated early during primary HIV-1 infection, had low viral diversity in both plasma or cells isolated after years on therapy (Josefsson et al., 2013; Kearney et al., 2014). Even though diversity values were low at the baseline visit, we detected a slight decline in HIV DNA diversity in PBMCs and a significant decline in proviral population complexity after 12 months on cART. As we found no evidence of ongoing viral replication, the reduction in intrahost proviral complexity following cART may be explained by the clearance of HIV-infected cells, probably long-lived T cells (Palmer et al., 2008), and coincided with a reduction in HIV DNA levels. Interestingly, the diversity indices exhibited by these subjects after 1 year of suppressive cART, were similar to those found in a subgroup of rare individuals capable of naturally controlling HIV-1 replication and maintaining it at low levels (named elite controllers). The diversity indices exhibited by elite controllers are associated with the presence of a putative more efficient mechanism to control HIV-1 replication and disease progression (de Azevedo et al., 2017). A previous study found an association between HIV genetic diversity during the early phase of infection and a faster HIV DNA decline following cART (Wang et al., 2017). In contrast, we did not find any significant correlation between immunological recovery or virologic control achieved after 1 year of suppressive cART and HIV diversity before cART initiation.

We detected three HIV-1 *env* subtypes (B, C, and F1) among the ten patients included in this study. This subtype distribution is in accordance with the molecular epidemiology found in Brazil (Bello et al., 2011). The different HIV-1 clades seem not to have influenced cART outcomes observed in the different subjects. However, it is noteworthy that IVA05 and IVA38, both infected with HIV-1 subtype C strains, presented the highest

values of viral diversity and complexity in both PBMC and plasma compartments. This finding agrees with a previous study showing that subtype C displays higher median diversity than subtype B (Rieder et al., 2011). Additionally, the majority of HIV-1 variants were R5-tropic during the entire follow-up period. The only exceptions were two subjects infected with HIV-1 subtype B, who exhibited a low frequency of X4-tropic clones even at the very beginning of HIV-1 infection, in agreement with previous studies (Sheppard et al., 2002; Raymond et al., 2010; Chalmet et al., 2012).

Despite the intensification of offering rapid tests for HIV detection inside and outside the health services, covering key populations at risk, and the ongoing follow-up of cohorts of serodiscordant couples conducted by our institution (INI-FIOCRUZ), the identification of individuals in the early primary infection phases, according to the Fiebig classification, and their follow-up are still hard tasks, as well as, the obtaining of high quality cryopreserved PMBC samples from long term storage to perform the laboratory analyses. Consequently, most of the studies addressing this group, considering the complex strategies for characterizing the early stages of HIV-acute infection, management of early cART starting and the clinical follow-up at specific time points limits the scope, mainly in developing countries where resources to closely following clinical cohorts are more restricted. Thus, the effects of early treatment initiation on the viral evolution, size and distribution of the viral reservoir are still poorly explored and need to be better characterized and, although having a relatively small number of individuals included in our study, we still justified conducting the present analyzes. As of our knowledge, this is the first Brazilian study discussing the viral evolution and reservoir size overtime in a cohort of HIV infected individuals starting cART in the early Fiebig stages of HIV primary infection. Further studies with larger cohorts and longer follow-up periods should be performed to identify better predictive markers for individuals, who might reach the PTC status after cART interruption. Assessment of HIV DNA level and diversity in different T cell subsets and/or compartments would substantially improve our understanding of the effect early cART has in these subjects and would likely add to the growing body of evidence indicating the contribution of each cell population to maintaining the HIV reservoir (Buzon et al., 2014).

In summary, we demonstrate that based on this small cohort and analyses of only the envelope region, 1 year of suppressive cART initiated in the early stages of HIV infection suggested a reduction of the size and complexity of HIV-1 total DNA in the PBMC reservoir, as well as achieve immunological restoration. Immunological recovery and virologic suppression were not associated with proviral diversity before cART onset. The early initiation of cART in HIV-acutely infected individuals in current Brazilian setting may favor strategies to achieve post-treatment control of HIV and, ultimately, a functional cure by restricting the pool of variants and allowing a more focused targeting with therapeutic vaccines or other immune approaches. However, further studies are required to determine the follow-up time necessary to achieve the lowest viral endpoints for clinical management.

AUTHOR CONTRIBUTIONS

MG, MM, VV, FC, and BG conceived and designed the study. TL, ED, and FC performed the experiments. AF, SC, and MA participated in patient recruitment. TL, ED, and MG analyzed the data and drafted the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

ED was financed by a Postdoctoral fellowship from the "Programa Nacional de Pós-Doutorado (PNPD)" by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001. MG and MM were recipients of a CNPq Fellowship, and TL was funded by a CAPES Ph.D. fellowship.

ACKNOWLEDGMENTS

We thank Soraia Moura and Sandro Nazer from Laboratório de Pesquisa Clínica em DST e AIDS for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00145/full#supplementary-material

REFERENCES

Abrahams, M.-R., Anderson, J. A., Giorgi, E. E., Seoighe, C., Mlisana, K., Ping, L.-H., et al. (2009). Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* 83, 3556–3567. doi: 10.1128/JVI.02132-08

Ananworanich, J., Chomont, N., Eller, L. A., Kroon, E., Tovanabutra, S., Bose, M., et al. (2016). HIV DNA set point is rapidly established in acute HIV infection and dramatically reduced by early ART. *EBioMedicine* 11, 68–72. doi: 10.1016/j.ebiom.2016.07.024

Archin, N. M., Vaidya, N. K., Kuruc, J. D., Liberty, A. L., Wiegand, A., Kearney, M. F., et al. (2012). Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 infection without accelerating the decay of latent infection. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9523–9528. doi: 10.1073/pnas.1120248109

Avettand-Fènoel, V., Hocqueloux, L., Ghosn, J., Cheret, A., Frange, P., Melard, A., et al. (2016). Total HIV-1 DNA, a marker of viral reservoir dynamics with clinical implications. *Clin. Microbiol. Rev.* 29, 859–880. doi: 10.1128/CMR.00015-16

Bar, K. J., Li, H., Chamberland, A., Tremblay, C., Routy, J. P., Grayson, T., et al. (2010). Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J. Virol.* 84, 6241–6247. doi: 10.1128/JVI.00077-10

Batorsky, R., Kearney, M. F., Palmer, S. E., Maldarelli, F., Rouzine, I. M., and Coffin, J. M. (2011). Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5661–5666. doi: 10.1073/pnas.1102036108

Bello, G., Soares, M. A., and Schrago, C. G. (2011). The Use of bioinformatics for studying HIV evolutionary and epidemiological history in South America. *AIDS Res. Treat.* 2011, 1–13. doi: 10.1155/2011/154945

Brodin, J., Zanini, F., Thebo, L., Lanz, C., Bratt, G., Neher, R. A., et al. (2016). Establishment and stability of the latent HIV-1 DNA reservoir. *eLife* 5:e18889. doi: 10.7554/eLife.18889

Buzon, M. J., Martin-Gayo, E., Pereyra, F., Ouyang, Z., Sun, H., Li, J. Z., et al. (2014). Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. *J. Virol.* 88, 10056–10065. doi: 10.1128/JVI.01046-14

Buzón, M. J., Massanella, M., Llibre, J. M., Esteve, A., Dahl, V., Puertas, M. C., et al. (2010). HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nat. Med.* 16, 460–465. doi: 10.1038/nm.2111

Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., et al. (2014). Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 345:1254031. doi: 10.1126/science.1254031

Chalmet, K., Dauwe, K., Foquet, L., Baatz, F., Seguin-Devaux, C., Van Der Gucht, B., et al. (2012). Presence of CXCR4-using HIV-1 in patients with recently diagnosed infection: correlates and evidence for transmission. *J. Infect. Dis.* 205, 174–184. doi: 10.1093/infdis/jir714

Chéret, A., Nembot, G., Mélard, A., Lascoux, C., Slama, L., Miailhes, P., et al. (2015). Intensive five-drug antiretroviral therapy regimen versus standard triple-drug therapy during primary HIV-1 infection (OPTIPRIM-ANRS 147): a randomised, open-label, phase 3 trial. *Lancet Infect. Dis.* 15, 387–396. doi: 10.1016/S1473-3099(15)70021-6

Chomont, N., DaFonsecaa, S., Vandergeetena, C., Ancuta, P., and Sékaly, R. P. (2011). Maintenance of CD4+ T-cell memory and HIV persistence: keeping memory, keeping HIV. *Curr. Opin. HIV AIDS* 6, 30–36. doi: 10.1097/COH.0b013e3283413775

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F. A., Yassine-Diab, B., et al. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* 15, 893–900. doi: 10.1038/nm.1972

Chun, T.-W., Justement, J. S., Pandya, P., Hallahan, C. W., McLaughlin, M., Liu, S., et al. (2002). Relationship between the size of the human immunodeficiency virus type 1 (HIV-1) reservoir in peripheral blood CD4 + T cells and CD4 +:CD8 + T cell ratios in aviremic HIV-1–infected individuals receiving long-term highly active antiretroviral therapy. *J. Infect. Dis.* 185, 1672–1676. doi: 10.1086/340521

Chun, T.-W., Stuyver, L., Mizell, S. B., Ehler, L. A., Mican, J. A. M., Baseler, M., et al. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* 94, 13193–13197. doi: 10.1073/pnas.94.24.13193

Colby, D. J., Trautmann, L., Pinyakorn, S., Leyre, L., Pagliuzza, A., Kroon, E., et al. (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig i acute HIV infection brief-communication. *Nat. Med.* 24, 923–926. doi: 10.1038/s41591-018-0026-6

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

de Azevedo, S. S. D., Caetano, D. G., Côrtes, F. H., Teixeira, S. L. M., dos Santos Silva, K., Hoagland, B., et al. (2017). Highly divergent patterns of genetic diversity and evolution in proviral quasispecies from HIV controllers. *Retrovirology* 14:29. doi: 10.1186/s12977-017-0354-5

Delwart, E., Magierowska, M., Royz, M., Foley, B., Peddada, L., Smith, R., et al. (2002). Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. *AIDS* 16, 189–195. doi: 10.1097/00002030-200201250-00007

Delwart, E. L., Shpaer, E. G., Louwagie, J., McCutchan, F. E., Grez, M., Rübsamen-Waigmann, H., et al. (1993). Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science* 262, 1257–1261. doi: 10.1126/science.8235655

Douek, D. C., Brenchley, J. M., Betts, M. R., Ambrozak, D. R., Hill, B. J., Okamoto, Y., et al. (2002). HIV preferentially infects HIV-specific CD4+ T cells. *Nature* 417, 95–98. doi: 10.1038/417095a

Ferreira, A. C. G., Coelho, L. E., Grinsztejn, E., Jesus, C. S. D., Guimarães, M. L., Veloso, V. G., et al. (2017). Transmitted drug resistance in patients with acute/recent HIV infection in Brazil. *Braz. J. Infect. Dis.* 21, 396–401. doi: 10.1016/j.bjid.2017.03.013

Fiebig, E. W., Wright, D. J., Rawal, B. D., Garrett, P. E., Schumacher, R. T., Peddada, L., et al. (2003). Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* 17, 1871–1879. doi: 10.1097/01.aids.0000076308.76477.b8

Ghosn, J., Taiwo, B., Seedat, S., Autran, B., and Katlama, C. (2018). HIV. *Lancet* 392, 685–697. doi: 10.1016/S0140-6736(18)31311-4

Gottlieb, G. S., Heath, L., Nickle, D. C., Wong, K. G., Leach, S. E., Jacobs, B., et al. (2008). HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. *J. Infect. Dis.* 197, 1011–1015. doi: 10.1086/529206

Gregori, J., Perales, C., Rodriguez-Frias, F., Esteban, J. I., Quer, J., and Domingo, E. (2016). Viral quasispecies complexity measures. *Virology* 493, 227–237. doi: 10.1016/j.virol.2016.03.017

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and 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

Hayashida, T., Tsuchiya, K., Kikuchi, Y., Oka, S., and Gatanaga, H. (2017). Emergence of CXCR4-tropic HIV-1 variants followed by rapid disease progression in hemophiliac slow progressors. *PLoS One* 12:e0177033. doi: 10.1371/journal.pone.0177033

Hecht, F. M., Wang, L., Collier, A., Little, S., Markowitz, M., Margolick, J., et al. (2006). A multicenter observational study of the potential benefits of initiating combination antiretroviral therapy during acute HIV infection. *J. Infect. Dis.* 194, 725–733. doi: 10.1086/506616

Hocqueloux, L., Avettand-Fènoël, V., Jacquot, S., Prazuck, T., Legac, E., Mélard, A., et al. (2013). Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. *J. Antimicrob. Chemother.* 68, 1169–1178. doi: 10.1093/jac/dks533

Hoenigl, M., Chaillon, A., and Little, S. J. (2016). CD4/CD8 cell ratio in acute HIV infection and the impact of early antiretroviral therapy: table 1. *Clin. Infect. Dis.* 63, 425–426. doi: 10.1093/cid/ciw293

Josefsson, L., von Stockenstrom, S., Faria, N. R., Sinclair, E., Bacchetti, P., Killian, M., et al. (2013). The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4987–E4996. doi: 10.1073/pnas.1308313110

Kearney, M., Maldarelli, F., Shao, W., Margolick, J. B., Daar, E. S., Mellors, J. W., et al. (2009). Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *J. Virol.* 83, 2715–2727. doi: 10.1128/JVI.01960-08

Kearney, M. F., Spindler, J., Shao, W., Yu, S., Anderson, E. M., O'Shea, A., et al. (2014). Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog.* 10:e1004010. doi: 10.1371/journal.ppat.1004010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kiselinova, M., Geretti, A. M., Malatinkova, E., Vervisch, K., Beloukas, A., Messiaen, P., et al. (2015). HIV-1 RNA and HIV-1 DNA persistence during suppressive ART with PI-based or nevirapine-based regimens. *J. Antimicrob. Chemother.* 70, 3311–3316. doi: 10.1093/jac/dkv250

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lengauer, T., Sander, O., Sierra, S., Thielen, A., and Kaiser, R. (2007). Bioinformatics prediction of HIV coreceptor usage. *Nat. Biotechnol.* 25, 1407–1410. doi: 10.1038/nbt1371

Li, H., Bar, K. J., Wang, S., Decker, J. M., Chen, Y., Sun, C., et al. (2010). High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathog.* 6:e1000890. doi: 10.1371/journal.ppat.1000890

Lorenzo-Redondo, R., Fryer, H. R., Bedford, T., Kim, E. Y., Archer, J., Kosakovsky Pond, S. L., et al. (2016). Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* 530, 51–56. doi: 10.1038/nature16933

Lori, F., Jessen, H., Lieberman, J., Finzi, D., Rosenberg, E., Tinelli, C., et al. (1999). Treatment of human immunodeficiency virus infection with Hydroxyurea, Didanosine, and a Protease inhibitor before seroconversion is associated with normalized immune parameters and limited viral reservoir. *J. Infect. Dis.* 180, 1827–1832. doi: 10.1086/315113

Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I. L., et al. (2014). Challenges in RNA virus bioinformatics. *Bioinformatics* 30, 1793–1799. doi: 10.1093/bioinformatics/btu105

Ngo-Giang-Huong, N., Deveau, C., Da Silva, I., Pellegrin, I., Venet, A., Harzic, M., et al. (2001). Proviral HIV-1 DNA in subjects followed since primary HIV-1 infection who suppress plasma viral load after one year of highly active antiretroviral therapy. *AIDS* 15, 665–673. doi: 10.1097/00002030-200104130-00001

Novitsky, V., Wang, R., Margolin, L., Baca, J., Rossenkhan, R., Moyo, S., et al. (2011). Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. *PLoS One* 6:e16714. doi: 10.1371/journal.pone.0016714

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017). *Vegan: Community Ecology Package. R Package Version. 2.0-10.*

Palmer, S., Kearney, M., Maldarelli, F., Halvas, E. K., Bixby, C. J., Bazmi, H., et al. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 43, 406–413. doi: 10.1128/JCM.43.1.406-413.2005

Palmer, S., Maldarelli, F., Wiegand, A., Bernstein, B., Hanna, G. J., Brun, S. C., et al. (2008). Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3879–3884. doi: 10.1073/pnas.0800050105

Piketty, C., Weiss, L., Assoumou, L., Burgard, M., Mélard, A., Ragnaud, J. M., et al. (2010). A high HIV DNA level in PBMCs at antiretroviral treatment interruption predicts a shorter time to treatment resumption, independently of the CD4 nadir. *J. Med. Virol.* 82, 1819–1828. doi: 10.1002/jmv.21907

Pilcher, C. D., Eron, J. J., Galvin, S., Gay, C., and Cohen, M. S. (2004). Acute HIV revisited: new opportunities for treatment and prevention. *J. Clin. Invest.* 113, 937–945. doi: 10.1172/JCI21540

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Encinas, S., Souyris, C., et al. (2010). CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. *AIDS* 24, 2305–2312. doi: 10.1097/QAD.0b013e32833e50bb

Richman, D. D., Margolis, D. M., Delaney, M., Greene, W. C., Hazuda, D., and Pomerantz, R. J. (2009). The challenge of finding a cure for HIV infection. *Science* 323, 1304–1307. doi: 10.1126/science.1165706

Rieder, P., Joos, B., Scherrer, A. U., Kuster, H., Braun, D., Grube, C., et al. (2011). Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) diversity and tropism in 145 patients with primary HIV-1 infection. *Clin. Infect. Dis.* 53, 1271–1279. doi: 10.1093/cid/cir725

Rose, P. P., and Korber, B. T. (2000). Detecting hypermutations in viral sequences with an emphasis on G -> a hypermutation. *Bioinformatics* 16, 400–401. doi: 10.1093/bioinformatics/16.4.400

Sáez-Cirión, A., Bacchus, C., Hocqueloux, L., Avettand-Fenoel, V., Girault, I., Lecuroux, C., et al. (2013). Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI study. *PLoS Pathog.* 9:e1003211. doi: 10.1371/journal.ppat.1003211

Schiffer, V., Deveau, C., Meyer, L., Iraqui, I., Nguyen-Wartel, A., Chaix, M.-L., et al. (2004). Recent changes in the management of primary HIV-1 infection: results from the French PRIMO cohort. *HIV Med.* 5, 326–333. doi: 10.1111/j.1468-1293.2004.00231.x

Sheppard, H. W., Celum, C., Michael, N. L., O'Brien, S., Dean, M., Carrington, M., et al. (2002). HIV-1 infection in individuals with the CCR5-Δ32/Δ32 genotype: acquisition of syncytium-inducing virus at seroconversion. *JAIDS J. Acquir. Immune Defic. Syndr.* 29, 307–313. doi: 10.1097/00126334-200203010-00013

Strain, M. C., Little, S. J., Daar, E. S., Havlir, D. V., Günthard, H. F., Lam, R. Y., et al. (2005). Effect of treatment, during primary infection, on establishment and clearance of cellular reservoirs of HIV-1. *J. Infect. Dis.* 191, 1410–1418. doi: 10.1086/428777

Van Zyl, G. U., Katusiime, M. G., Wiegand, A., McManus, W. R., Bale, M. J., Halvas, E. K., et al. (2017). No evidence of HIV replication in children on antiretroviral therapy. *J. Clin. Invest.* 127, 3827–3834. doi: 10.1172/JCI94582

von Stockenstrom, S., Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., et al. (2015). Longitudinal genetic characterization reveals that cell proliferation maintains a persistent HIV type 1 DNA pool during effective HIV therapy. *J. Infect. Dis.* 212, 596–607. doi: 10.1093/infdis/jiv092

Wang, N., Li, Y., Han, Y., Xie, J., and Li, T. (2017). HIV sequence diversity during the early phase of infection is associated with HIV DNA reductions during antiretroviral therapy. *J. Med. Virol.* 89, 982–988. doi: 10.1002/jmv.24723

Yerly, S., Kaiser, L., Perneger, T. V., Cone, R. W., Opravil, M., Chave, J. P., et al. (2000). Time of initiation of antiretroviral therapy: impact on HIV-1 viraemia. The Swiss HIV cohort study. *AIDS* 14, 243–249. doi: 10.1097/00002030-200002180-00006

Yukl, S. A., Shergill, A. K., McQuaid, K., Gianella, S., Lampiris, H., Hare, C. B., et al. (2010). Effect of raltegravir-containing intensification on HIV burden and T-cell activation in multiple gut sites of HIV-positive adults on suppressive antiretroviral therapy. *AIDS* 24, 2451–2460. doi: 10.1097/QAD.0b013e32833ef7bb

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 February 2019
doi: 10.3389/fmicb.2019.00097

Identification of New HIV-1 Circulating Recombinant Forms CRF81_cpx and CRF99_BF1 in Central Western Brazil and of Unique BF1 Recombinant Forms

*Mônica N. G. Reis[1], Monick L. Guimarães[2], Gonzalo Bello[2] and Mariane M. A. Stefani[1]**

[1] Laboratório de Imunologia da Aids e da Hanseníase, Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, Brazil, [2] Laboratório de Aids e Imunologia Molecular, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:
Michael M. Thomson, Instituto de Salud Carlos III, Spain

Reviewed by:
Timokratis Karamitros, University of Oxford, United Kingdom
Antoinette Van Der Kuyl, University of Amsterdam, Netherlands

***Correspondence:**
Mariane M. A. Stefani
mmastefani@gmail.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *12 October 2018*
Accepted: *16 January 2019*
Published: *11 February 2019*

Citation:
Reis MNG, Guimarães ML, Bello G and Stefani MMA (2019) Identification of New HIV-1 Circulating Recombinant Forms CRF81_cpx and CRF99_BF1 in Central Western Brazil and of Unique BF1 Recombinant Forms. Front. Microbiol. 10:97. doi: 10.3389/fmicb.2019.00097

Intersubtype recombinants classified as circulating recombinant forms (CRFs) or unique recombinant forms (URFs) have been shown to play an important role in the complex and dynamic Brazilian HIV/AIDS epidemic. Previous *pol* region studies (2003–2013) in 828 patients from six states from Central Western, Northern and Northeastern Brazil reported variable rates of BF1, F1CB, BC, and CF1 mosaics. In this study HIV-1 subtype diversity BF1, F1CB, BC, and CF1 recombinants in *pol* were analyzed. Full/near-full/partial genome sequences were generated from F1CB and BF1 recombinants. Genomic DNA extracted from whole blood was used in nested-PCR to amplify four overlapping fragments encompassing the full HIV-1 genome. Phylogenetic trees were generated using the neighbor-joining/NJ method (MEGA 6.0). The time of the most recent common ancestor (TMRCA) of F1CB and BF1 clades was estimated using a Bayesian Markov Chain Monte Carlo approach (BEAST v1.8; BEAGLE). Bootscanning was used for recombination analyses (Simplot v3.5.1); separate NJ phylogenetic analysis of fragments confirmed subtypes. The phylogenetic analyses of protease/reverse-transcriptase sequences in 828 patients revealed 76% subtype B (n = 629), 6.4% subtype C (n = 53), 4.2% subtype F1 (n = 35), 13.4% intersubtype recombinants: 10.5% BF1 (n = 87), 2.3% BC (n = 19), 0.4% F1CB (n = 3), and 0.2% CF1 (n = 2). Two full and one partial BF1C genomes allowed the characterization of the CRF81_cpx that has 9 breakpoints dividing the genome into 10 subregions. Basic Local Alignment Search Tool searches (Los Alamos HIV Sequence Database) identified six other sequences with the same recombination profile in *pol*, five from Brazil, and one from Italy. The estimated median TMRCA of CRF81_cpx was 1999 (1992–2003). CRF60_BC-like sequences, originally described in Italy, were also found. Two full and one near full-length BF1 genomes led to the characterization of the new CRF99_BF1 that has six recombination breakpoints dividing the genome into seven subregions. Two new URFs BF1, with six recombination breakpoints and seven subregions were

also characterized. The description of the first Brazilian BF1C CRF81_cpx and of the new CRF99_BF1 corroborate the important role of CRFs in the HIV/AIDS epidemic throughout Brazil. Our data also highlight the value of HIV-1 full-genome sequence studies in order to fully reveal the complexity of the epidemic in a huge country as Brazil.

Keywords: HIV-1, CRFs, URFs, molecular epidemiology, Central Western, Brazil

INTRODUCTION

Recombination of HIV-1 group M subtypes is considered a driving force for its genetic diversity worldwide (Worobey and Holmes, 1999). Brazil is a vast country with remarkable socio-economic and geographic differences where a very complex and dynamic HIV/AIDS epidemic has been reported. HIV-1 subtype B predominates in all geographic regions, except in Southern region where subtype C is a major variant (Soares et al., 2005; Gräf and Pinto, 2013; Delatorre et al., 2017). The distribution of F1 subtype has also significant regional variances ranging from low prevalence, in most regions, to high frequency in Pernambuco state, Northeastern Brazil (Cavalcanti et al., 2012; Gaspareto et al., 2012; Silveira et al., 2012; Lima et al., 2016a). The cocirculation of distinct HIV-1 subtypes is known to increase the chances of coinfection and of the generation of BF1, F1CB, BC, and CF1 intersubtype recombinant viruses (Jetzt et al., 2000; Faria et al., 2014). In the last decade, different circulating recombinant forms (CRFs) and unique recombinant forms (URFs) have been described including BC recombinants especially in the Southern and BF1 recombinants in different Brazilian regions (Guimarães et al., 2008, 2010; Passaes et al., 2009; Ferreira et al., 2011; da Silveira et al., 2012; Pessôa et al., 2014b, 2016; Moura et al., 2015a,b; Gräf et al., 2016; Lima et al., 2016b; Reis et al., 2017).

Previous studies from our group in the HIV-1 *pol* region reported variable rates of intersubtype recombinant BF1, F1CB, BC, and CF1 viruses among over 800 patients living in six Brazilian States located in the Central Western, Northern, and Northeastern Brazilian regions (Cardoso et al., 2009, 2010, 2011; Cardoso and Stefani, 2009; Carvalho et al., 2011; Ferreira et al., 2011; Alcântara et al., 2012; da Silveira et al., 2012; da Costa et al., 2013; Moura et al., 2015a,b; Lima et al., 2016b). Recently near full and full-length genomes revealed a noteworthy diversity of BF1 recombinants among these patients and allowed the identification of the CRF90_BF1 and of several URFs BF1 (Reis et al., 2017). In this study, we summarize HIV-1 molecular data in *pol* among this large group of patients from six Brazilian states. Additionally, we describe the frequency and profile of intersubtype recombinants. Full and near-full length genome sequences of F1CB and BF1 intersubtype mosaics allowed the characterization of two new CRFs identified in Central Western Brazil and partial genomes characterized two new URFs BF1.

MATERIALS AND METHODS

Study Samples

Our study population included 828 HIV-1 infected individuals recruited in Central Western (Goiás, Mato Grosso, and Mato Grosso do Sul states), Northern (Tocantins state), and Northeastern (Maranhão and Piauí states) Brazilian regions from 2003 to 2013. Patients belonged to the following groups: recently diagnosed antiretroviral/ARV-naïve (701/828; 84.7%), therapeutic failure of ARV treatment (127/828; 15.3%), inmates (27/828; 3.3%), and pregnant women (278/828; 33.6%). These studies were approved by the local Institutional Ethics Committee and all participants signed an informed consent form before blood collection (Goiás: protocols #073/05, #003/2008, #163/2010, at CEPMHA/HC/UFG; Mato Grosso: protocol #435/07 at Universidade Federal do Mato Grosso/UFMT; Mato Grosso do Sul: protocol #1143 at Universidade Federal do Mato Grosso do Sul/UFMS; Piauí: protocol #022/2011 at Universidade Estadual do Piauí/UESPI; Maranhão: protocol #16/2011 at Hospital de Doenças Tropicais Dr Natan Portela).

POL Gene Sequencing

Plasma RNA extraction, reverse transcription into complementary DNA (cDNA), amplification by nested polymerase chain reaction, sequencing of the protease (PR) and reverse transcriptase (RT) regions in the *pol* gene and phylogenetic analyses were previously described (Cardoso et al., 2009, 2010, 2011; Cardoso and Stefani, 2009; Carvalho et al., 2011; Ferreira et al., 2011; Alcântara et al., 2012; da Silveira et al., 2012; da Costa et al., 2013; Moura et al., 2015a,b; Lima et al., 2016b).

Full-Length, Near-Full-Length, and Partial HIV-1 Genomes

Genomic DNA was extracted from whole blood samples (QIAamp® DNA Blood Mini Kit/QIAGEN, Qiagen, Hilden, Germany). The double-stranded proviral HIV-1 DNA was amplified by nested-PCR employing Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, United States) into four overlapping fragments [(#1 (408–2594), #2 (2253–4830), #3 (4653–7811), and #4 (6954–9625) relative position to HXB2 genome] using HIV-1 specific primers as described in Reis et al. (2017). The amplified DNA fragments were purified (kit QIAquick® PCR Purification Kit/QIAGEN, Qiagen, Hilden, Germany) and sequenced (*Big Dye Terminator Sequencing Kit*

v. 3.1, Applied Biosystems, Foster City, CA, United States; ABI Prism 3100 Genetic Analyzer, Applied Biosystems, Foster City, CA, United States).

Phylogenetic and Recombination Analyses

Sequences from this study were aligned using Clustal X 2.0 implemented in BioEdit 7.2.0 program (Thompson et al., 1994). Reference sequences of HIV-1 group M subtypes (B, C, and F1) were obtained from the Los Alamos HIV Database (http://hiv.lanl.gov, last accessed December 2018). Phylogenetic trees were generated using the neighbor-joining (NJ) method (Nei and Kumar, 2002) under the Kimura two-parameter model (Kimura, 1980) using MEGA 6.0 software (Tamura et al., 2013). Bootstrap values (BP, 1000 replicates) above 80% were considered significant. Recombination analyses were performed in all viral isolates using the bootscanning method implemented in Simplot v3.5.1 software (Lole et al., 1999) with the following parameters: 300 nucleotide (nt) window, 20 nt increments, NJ method under Kimura's two-parameter correction with 100 bootstrap replicates. Informative site analyses were used to characterize the recombination breakpoints, for this purpose, consensus sequences from Brazilian HIV-1 subtypes B, C, and F1 were generated in the DAMBE program (Xia and Xie, 2001). The fragments of our study sequences assigned to specific HIV-1 subtypes were confirmed by separate NJ phylogenetic analysis as described above.

Representative samples from the HIV-1 BF1 and F1CB Brazilian clusters herein identified were submitted to a Basic Local Alignment Search Tool (BLAST) analysis in order to recover other sequences with high similarity (>95%) and probably with similar recombination profiles. The BLAST analysis was performed using sequences obtained from the Los

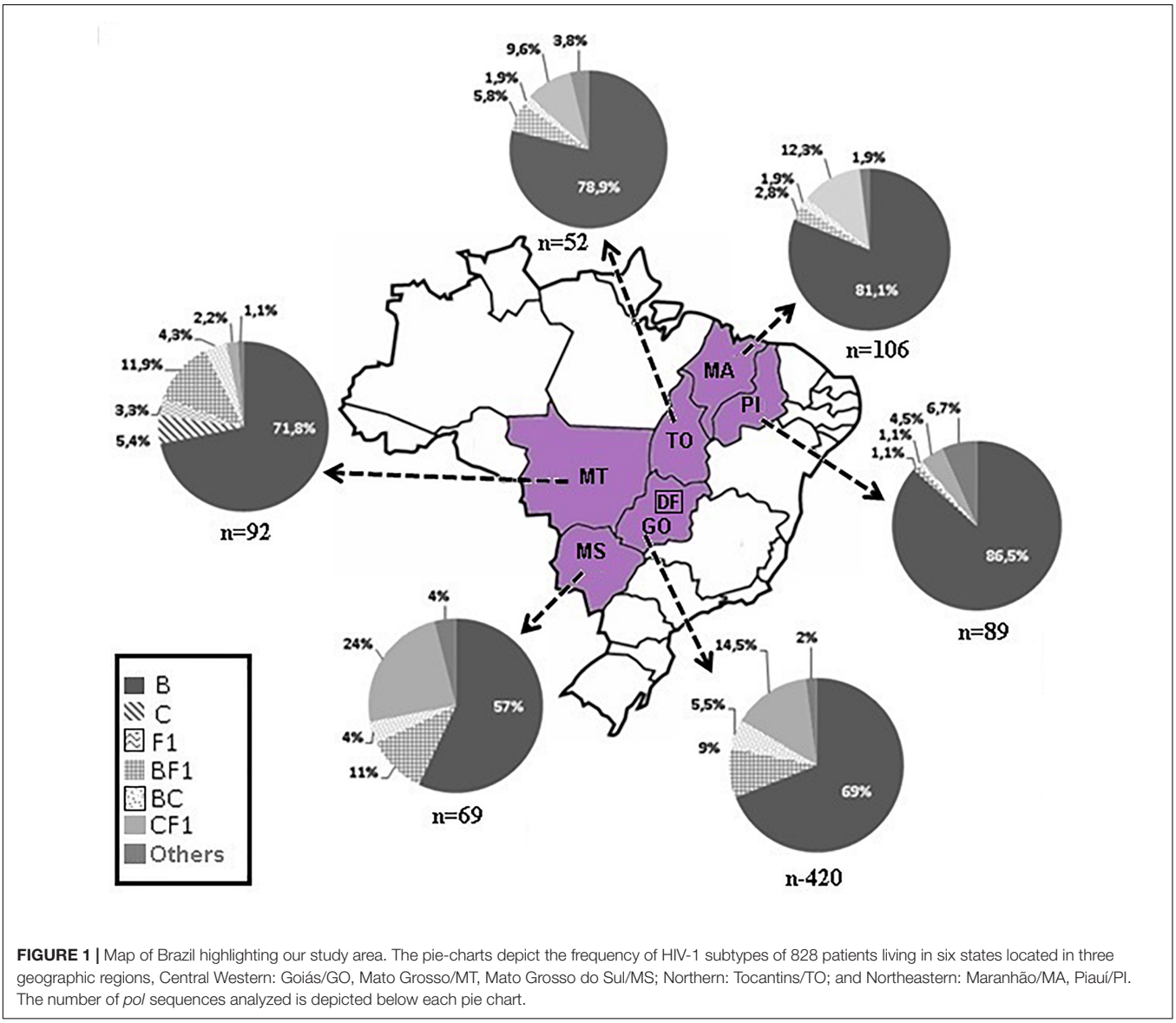


FIGURE 1 | Map of Brazil highlighting our study area. The pie-charts depict the frequency of HIV-1 subtypes of 828 patients living in six states located in three geographic regions, Central Western: Goiás/GO, Mato Grosso/MT, Mato Grosso do Sul/MS; Northern: Tocantins/TO; and Northeastern: Maranhão/MA, Piauí/PI. The number of *pol* sequences analyzed is depicted below each pie chart.

Alamos HIV Sequence Database (http://hiv.lanl.gov, last accessed December 2018).

Evolutionary Analyses of F1CB and BF1 Recombinants

The time of the most recent common ancestor (TMRCA) of HIV-1 F1CB and BF1 clades was estimated using a Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.8 (Drummond et al., 2002; Drummond and Rambaut, 2007) with BEAGLE to improve run-time (Suchard and Rambaut, 2009). Analyses were performed using the GTR+I+G nucleotide substitution model selected by the jModeltest program (Posada, 2008), a flexible Bayesian Skyline coalescent tree prior that does not require strong prior assumptions of demographic history (Drummond et al., 2002) and a relaxed uncorrelated lognormal molecular clock model (Drummond et al., 2006) with an informative uniform prior interval ($1.5–2.5 \times 10^{-3}$ nucleotide substitutions per site per year) as the estimated coefficient of rate variation (mean = 0.30; 95% highest probability density/HPD values: 0.0004–0.65) indicated a significant

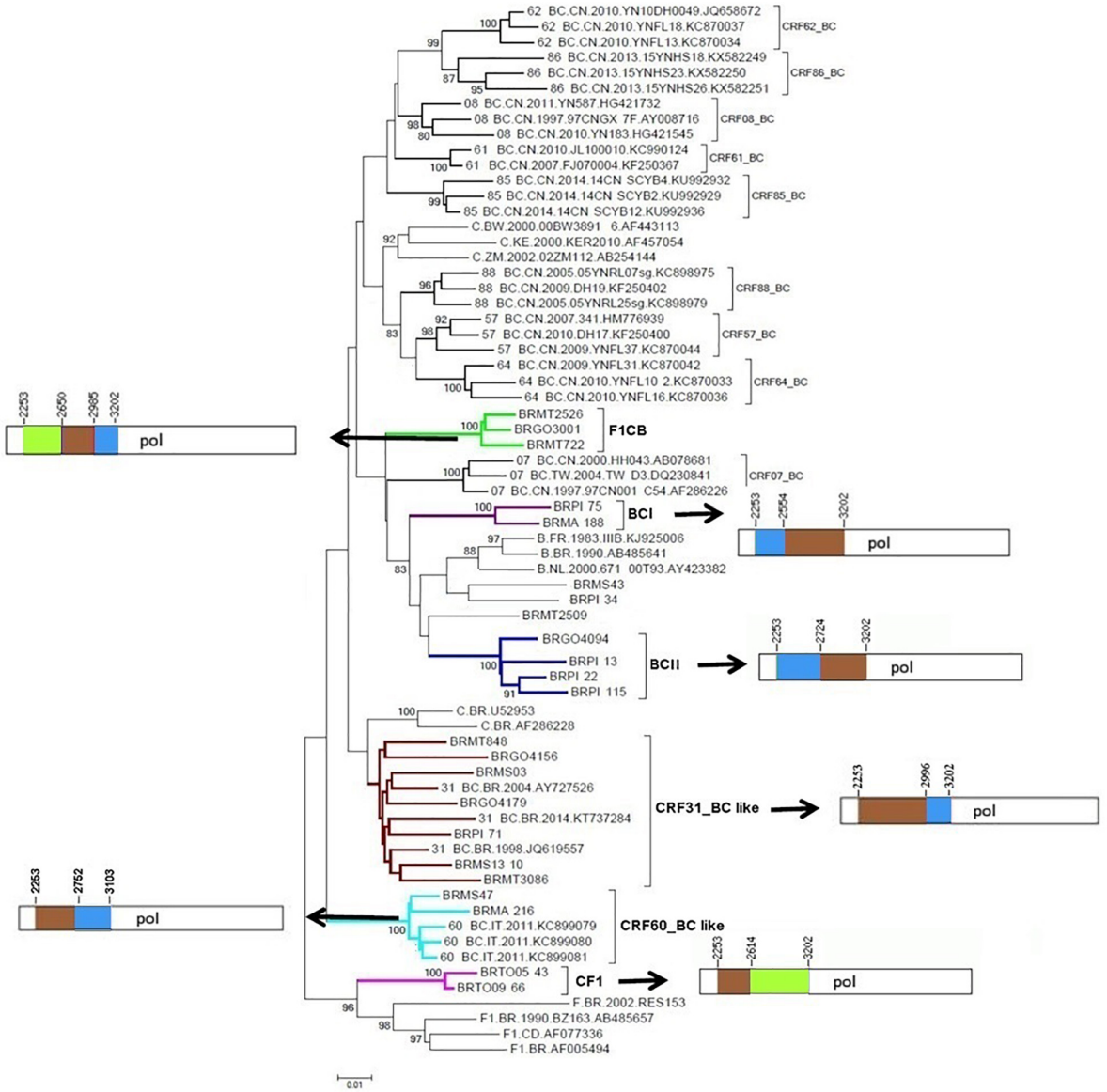


FIGURE 2 | Phylogenetic tree of partial HIV-1 *pol* sequences of F1CB, BC, and CF1 recombinants from Central Western, Northern, and Northeastern Brazil. The tree was constructed using MEGA software, 6.0 version under NJ method and Kimura two-parameter model (Bootstrap value over 80%). The mosaic patterns of recombinants: Cluster F1CB, Clusters BC (I and II), CRF31_BC-like, CRF60_BC-like, and Cluster CF1 are depicted. In the mosaic structure representations of F1CB, BC, and CF1 clusters, the breakpoint positions are indicated according to HXB2 genome position. In the mosaic structure, the green color stands for HIV-1 subtype F1, blue color stands for subtype B, and brown color stands for subtype C.

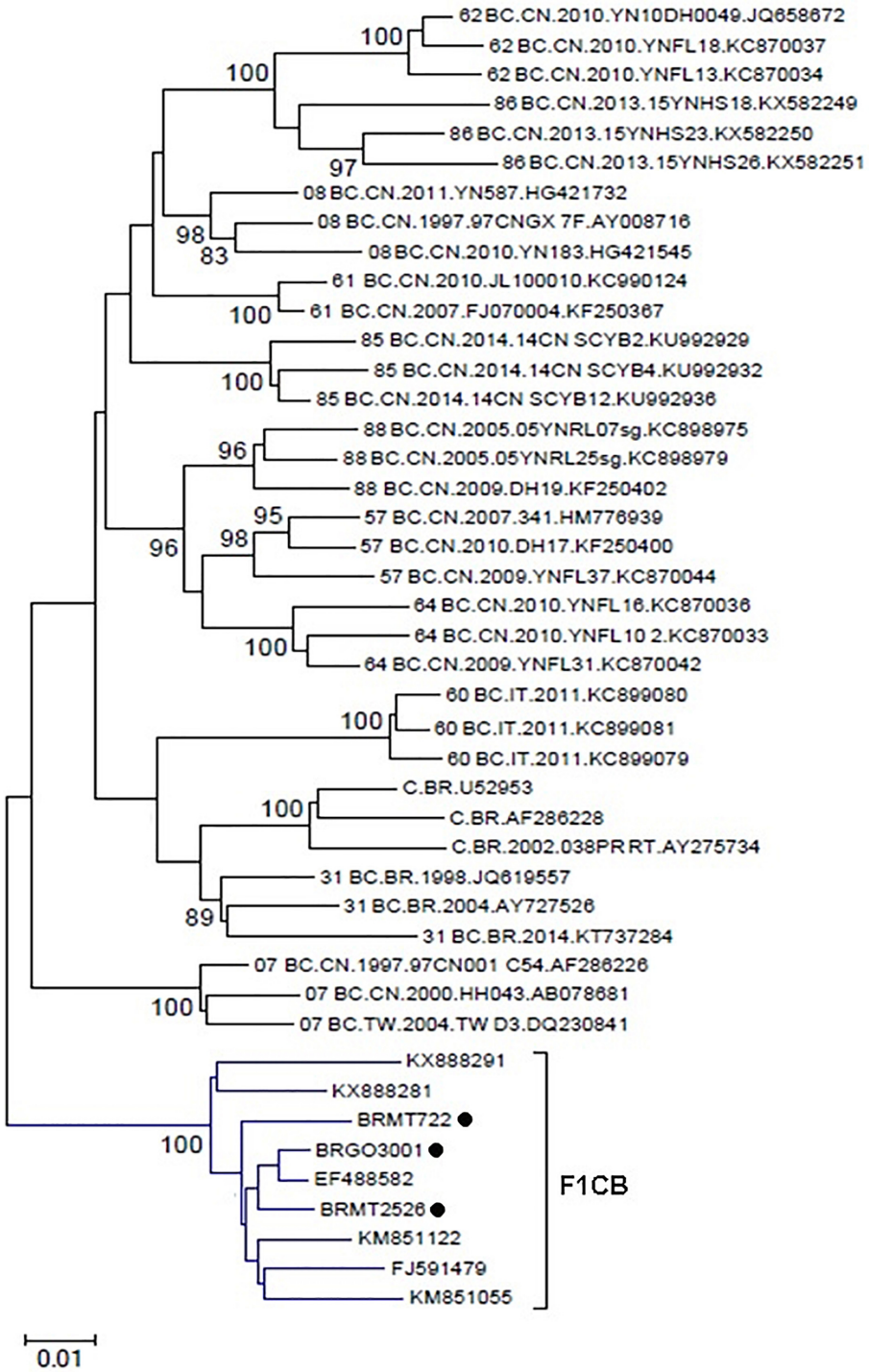


FIGURE 3 | Phylogenetic tree including the three F1CB sequences identified here and the six *pol* sequences sharing over 95% similarity recovered from the BLAST search. The tree was constructed using MEGA software, 6.0 version under NJ method and Kimura two-parameter model (Bootstrap value over 80%).

variation of substitution rate among branches. One MCMC chain was run for 1×10^7 generations. Convergence and uncertainty of parameter estimates were assessed by calculating the effective sample size (ESS) and the 95% HPD, respectively, using Tracer v1.6 (Rambaut and Drummond, 2007). The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.8 and visualized with FigTree v1.4.0.

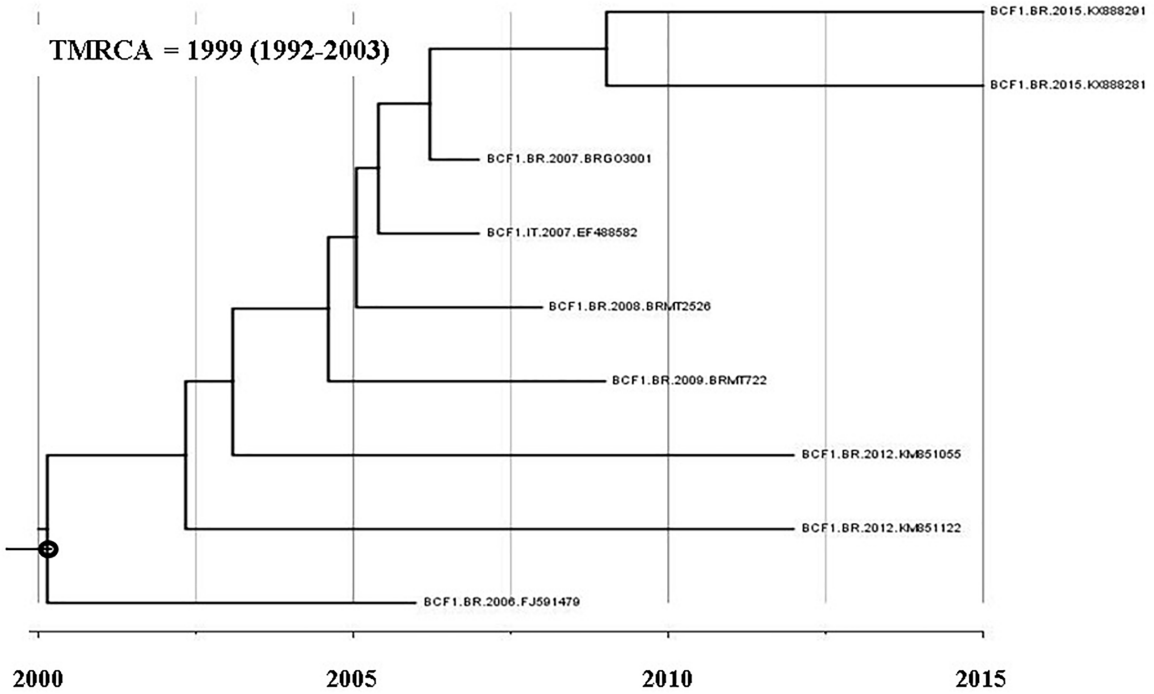


FIGURE 4 | Time-scaled Bayesian MCMC tree of nine *pol* sequences of F1CB HIV-1 sequences.

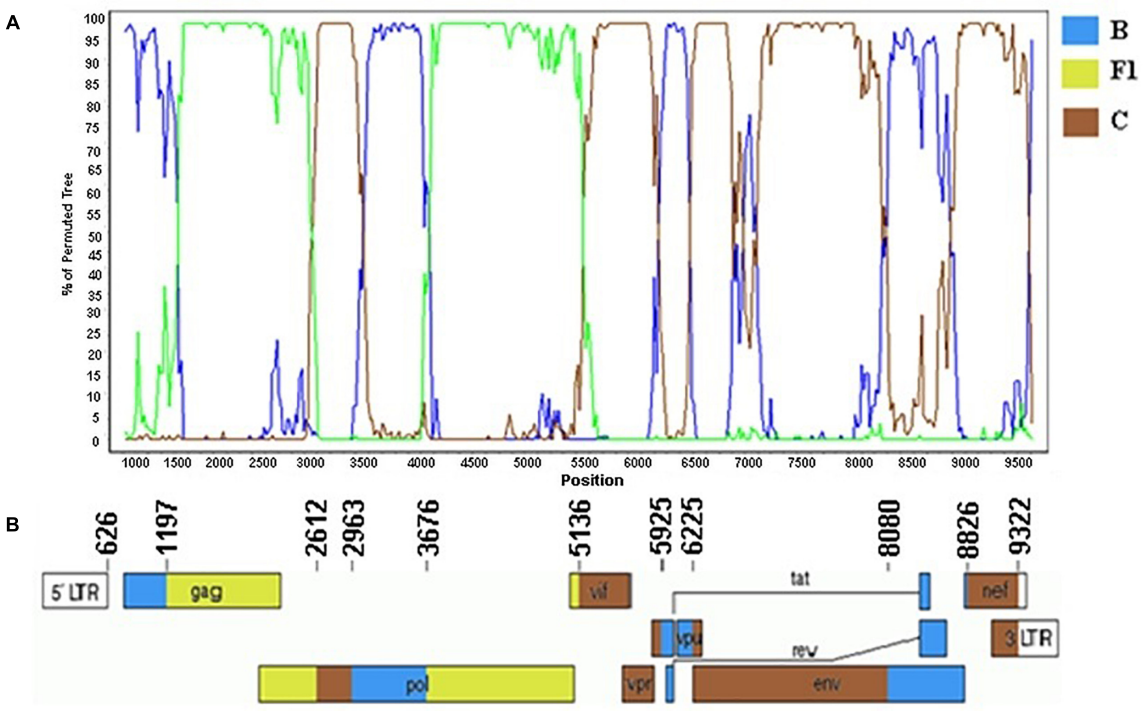


FIGURE 5 | Mosaic structure of the CRF81_cpx composed of HIV-1 subtypes B, F1, and C. **(A)** Bootscanning analysis was conducted using a window size of 300 bp and a step size of 20 bp along with reference strains of representative subtypes B, F1, and C from HIV-1 group M. **(B)** Genomic structure of CRF81_cpx. Breakpoint positions according to HXB2 genome positions are indicated. The blue color stands for HIV-1 subtype B, green color stands for subtype F1, and brown color stands for subtype C. The mosaic map was generated using the Recombinant HIV-1 Drawing Tool (https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html).

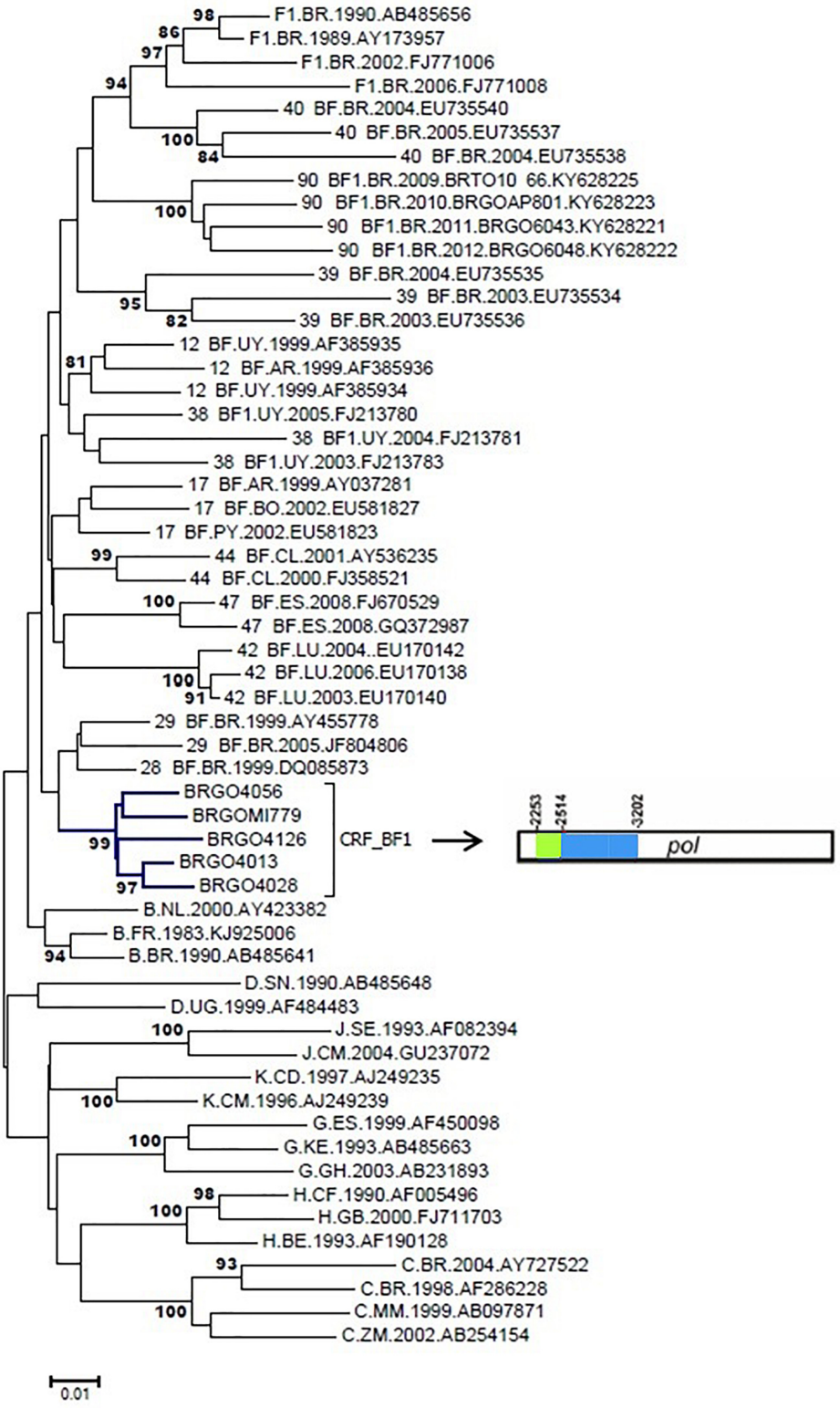


FIGURE 6 | Phylogenetic analysis of five *pol* BF1 sequences with similar recombination profile. NJ method and Kimura two-parameter evolutionary model/1,000 replicate *bootstrap* values were applied. In the mosaic structure, the green color stands for subtype F1 and the blue color stands for subtype B.

DATA AVAILABILITY

All HIV-1 sequences generated in this study were deposited in the GenBank database (accession numbers MH938677, MH938678, and MH986013-MH986018).

RESULTS

Phylogenetic Analyses of *pol* Region in HIV-1-Infected Patients From Six Brazilian States

The phylogenetic analyses of the *pol* (PR/RT) region in 828 patients from six Brazilian States showed that 76% was subtype B (629 out of 828), 6.4% was subtype C (53 out of 828), and 4.2% was subtype F1 (35 out of 828). Intersubtype recombinant viruses represented 13.4% (111 out of 828): 10.5% BF1 (87 out of 828), 2.3% BC (19 out of 828), 0.4% F1CB (3 out of 828), and 0.2% CF1 (2 out of 828) (**Figure 1**; **Supplementary Table 1**). The BF1 recombinants identified in this dataset were recently described in detail elsewhere (Reis et al., 2017).

HIV-1 F1CB, BC, and CF1 Recombinant Viruses in the *pol* Region

Phylogenetic and bootscanning analyses of the 23 mosaic sequences composed of subtypes F1CB, subtypes BC, and subtypes CF1 (**Supplementary Table 1**) showed six clusters with distinct recombination profiles (**Figure 2**). Three BC recombinant isolates (BRMS43, BRPI34, and BRMT2509) did not cluster with other samples.

The F1CB cluster was composed of three isolates (Goiás = 1, Mato Grosso = 2). The BC-I cluster contained two isolates (Maranhão = 1, Piauí = 1) and the BC-II cluster contained four isolates (Goiás n = 1, Piauí n = 3). The CRF31_BC-like cluster contained seven isolates (Goiás = 2, Mato Grosso do Sul = 2, Mato Grosso = 2, Piauí = 1). The CRF60_BC-like cluster comprised two isolates (Maranhão = 1, Mato Grosso do Sul = 1) and the CF1 cluster contained two isolates from Tocantins (**Figure 2**).

Characterization of the F1CB Cluster in the *pol* Region

The deduced mosaic structures of the three F1CB recombinant strains in *pol* (BRGO3001 from Goiás, BRMT2526 and BRMT722 from Mato Grosso) were compared and the patterns of the shared recombination breakpoints are summarized in **Figure 2**.

A BLAST search analysis recovered six other *pol* sequences with high homology with those from F1CB cluster. Five of them were from individuals collected in two Brazilian states (FJ591479, KM851055, and KM851122 from Paraná, Southern region and KX888281 and KX888291 from Mato Grosso). One recovered BF1C *pol* sequence was from Italy. Phylogenetic and recombination analyses of the *pol* F1CB sequences identified here and of these six sequences recovered from the BLAST search showed bootstrap values of 100% and the same mosaic structure

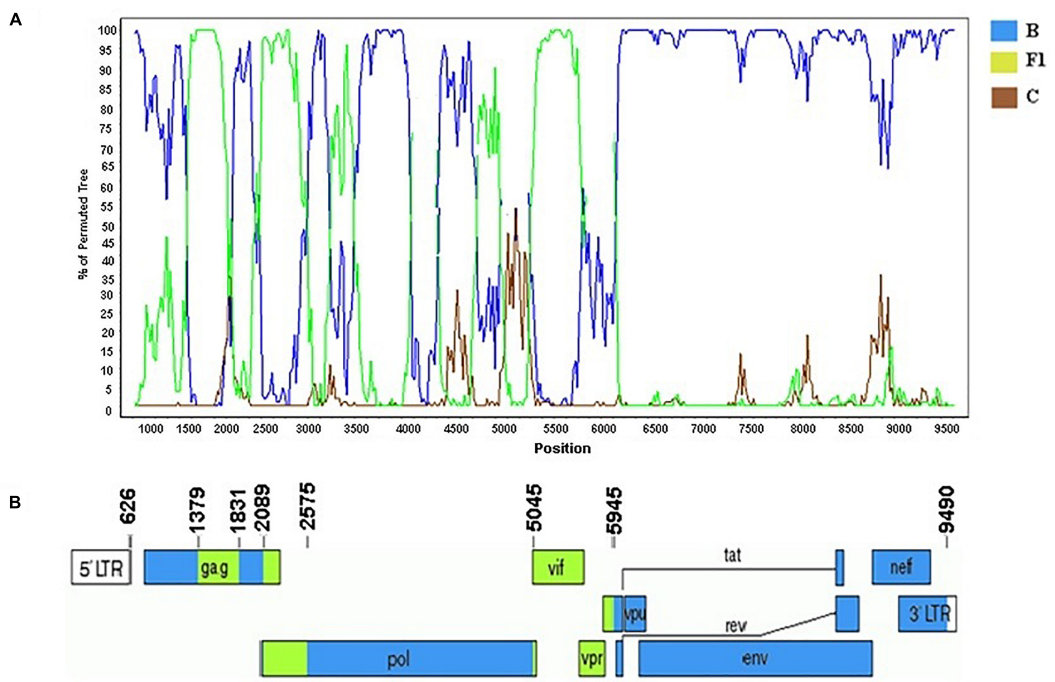


FIGURE 7 | Mosaic structure of the new CRF99_BF1. **(A)** Bootscanning analysis was conducted using a window size of 300 bp and a step size of 20 bp along with reference strains of B, C, and F1 representative HIV-1M subtypes. **(B)** Genomic structure of the CRF99_BF1. Breakpoint positions according to HXB2 genome numbering system are indicated. The blue color stands for subtype B, the green color stands for subtype F1, and the brown color for subtype C. The mosaic map was generated using the Recombinant HIV-1 Drawing Tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom mapper.html).

pattern (**Figure 3**). The estimated median TMRCA of the F1CB cluster was 1999 ranging from 1992 to 2003 (**Figure 4**).

Description of the CRF81_cpx

Phylogenetic and bootscanning analyses of the near-full and full-length genome sequences from the F1CB cluster: BRGO3001 strain (626–9322 bp), BRMT2526 strain (626–9322 bp), and partial sequences of the BRMT722 strain (626–1677 bp; 2263–4822 bp; 4824–5964 bp; 7855–8653 bp) described here allowed the description of the new CRF81_cpx named according to the standardized nomenclature by the Los Alamos National Laboratory.

The genome sequences of the three CRF81_cpx strains showed a complex recombination profile with nine breakpoints, which divide the HIV-1 genome into 10 subregions (**Figure 5**). The following subregions and subtypes were identified: subtype B (626–1196 bp; 2963–3675 bp; 5925–6224 bp; 8080–8825 bp), subtype F1 (1197–2611 bp; 3676–5135 bp), and subtype C (2612–2962 bp; 5136–5924 bp; 6225–8079 bp; 8826–9322 bp).

The three CRF81_cpx strains were obtained from two 40- and 46-year-old males who lived in Rondonópolis and Cuiabá cites, in Mato Grosso state and from one 33-year-old female living in Alexânia city, in Goiás state. All three patients were ARV-naïve and were blue-collar workers. The 46-year-old patient self-reported to belong to the men who have sex with men (MSM) category and the other two patients were from the heterosexual category.

Identification of the New CRF99_BF1

We have previously described five highly supported clusters (#1–5) among 87 BF1 recombinants and full-genome sequences from Cluster #5 characterized the new CRF90_BF1 (Reis et al., 2017; **Supplementary Table 1**).

In the current study five BF1 isolates from Cluster #2 (described in Reis et al., 2017) (BRGOMI779, BRGO4013, BRGO4028, BRGO4056, and BRGO4126) were further sequenced (**Figure 6**). Among these five BF1 isolates from Cluster #2 we were able to generate the full-length genome sequence of two viruses, the BRGO4056 strain (407–9616 bp) and the BRGO4028 strain (363–9612 bp) and the near-full-length genome sequence of BRGO4013 strain (350–7335 bp and 7750–8576 bp) relative to HXB2 genome positions.

According to the standardized nomenclature by the Los Alamos National Laboratory the full and near-full sequences of these BF1 strains were designated as CRF99_BF1 (BP = 99%). The mosaic structure of the CRF99_BF1 showed the predominance of subtype B (**Figure 7**). The bootscanning analysis of the whole genome of the CRF99_BF1 showed six recombination breakpoints which divide the genome into seven subregions (**Figure 7**). According to this, the following subregions and

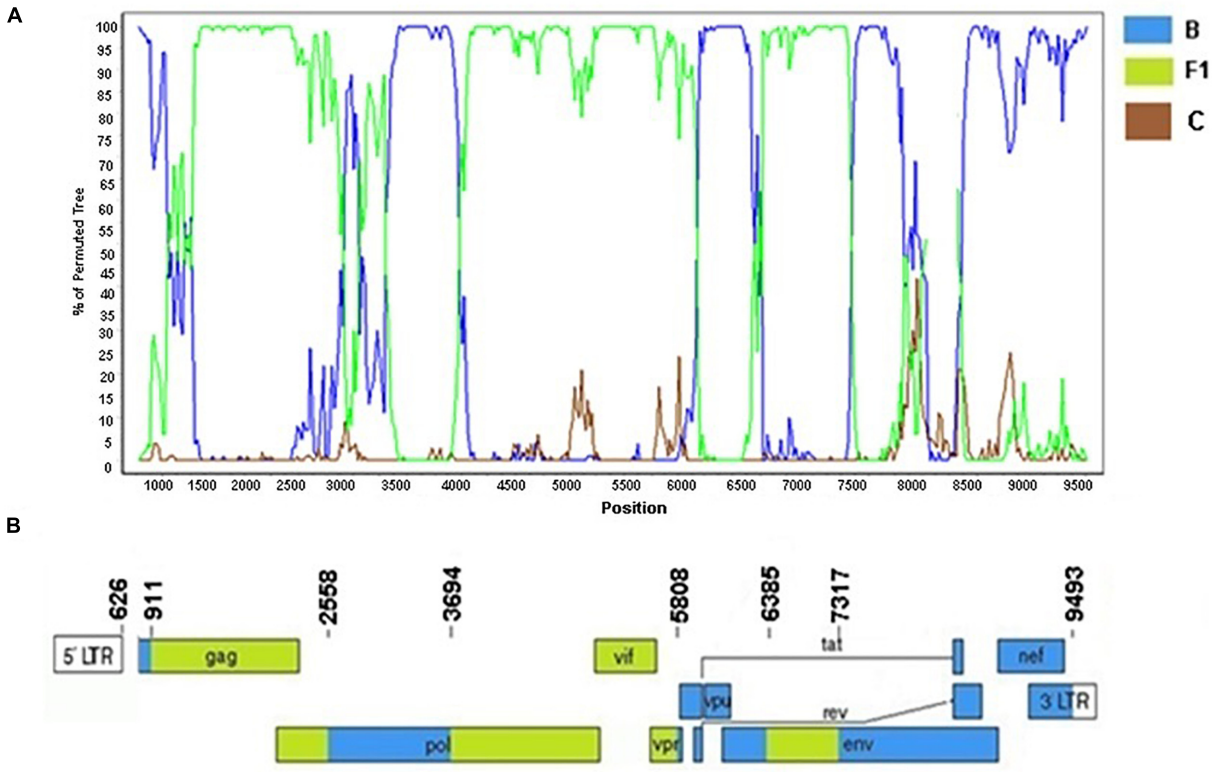


FIGURE 8 | Recombination breakpoint analyses of URF_BF1/BRGOMI744. **(A)** Bootscanning analysis was conducted using a window size of 300 bp and a step size of 20 bp along with reference strains of B, C, and F1 representative HIV-1 M subtypes. **(B)** Genomic structure of URF/BRGOMI744. The mosaic map was generated using the Recombinant HIV-1 Drawing Tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom mapper.html).

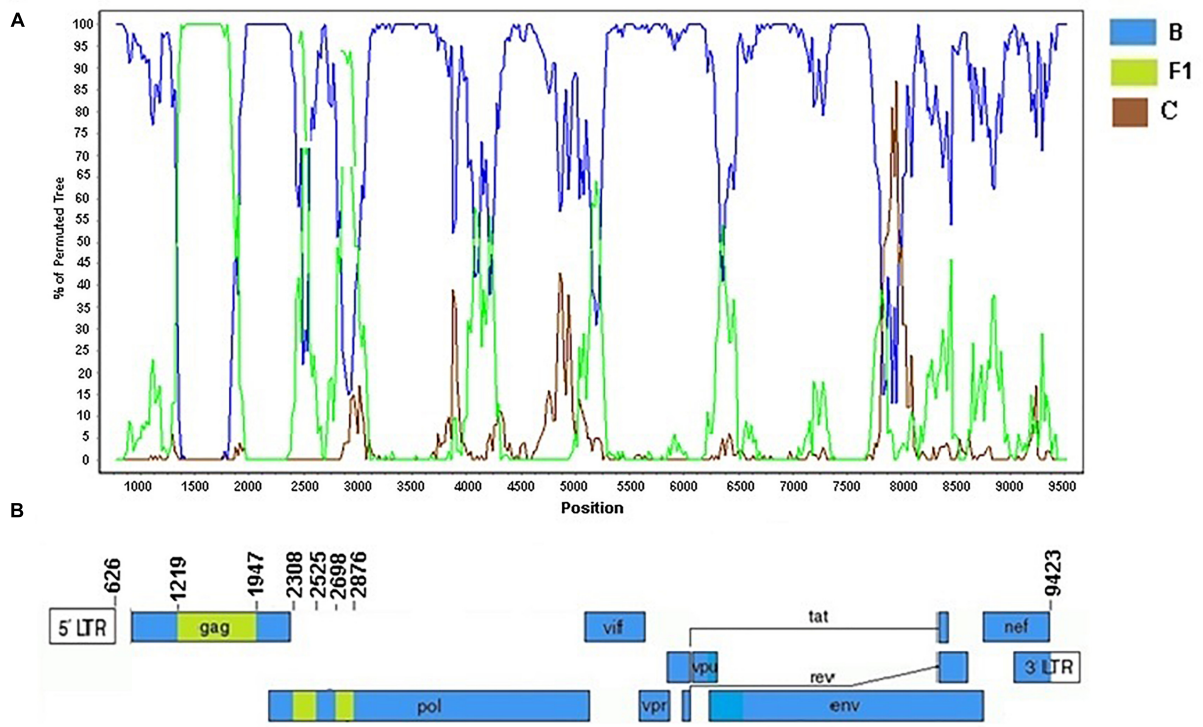


FIGURE 9 | Recombination breakpoint analyses of the URF_BF1 identified in the BRMT1319 strain. **(A)** Bootscanning analysis was conducted using a window size of 300 bp and a step size of 20 bp along with reference strains of B, C, and F1 representative HIV-1 M subtypes. **(B)** Genomic structure of the URF_BF1 identified in the BRMT1319 strain. The mosaic map was generated using the Recombinant HIV-1 Drawing Tool (www.hiv.lanl.gov/content/sequence/DRAW_CRF/recommapper.html).

subtypes were identified in the full-genome sequence of BRGO4056 strain: subtype B (626–1378 bp; 1831–2088 bp; 2575–5044 bp; and 5945–9601 bp) and subtype F1 (1379–1830 bp; 2089–2574 bp; and 5045–5944 bp) all of them relative to HXB2 genome positions.

The three strains that make up the CRF99_BF1 were obtained from pregnant women (17-, 25-, and 31-year-old) all of them ARV-naïve living in Goiás state and recruited from 2008 to 2009.

No other HIV-1 *pol* sequence sharing a mosaic profile similar to the sequences that make up the new CRF99_BF1 was identified in our BLAST searches. The estimated median TMRCA of the CRF99_BF1 was previously reported to be 1993, ranging from 1985 to 1998 (Reis et al., 2017).

Characterization of Two New URFs_BF1

Among six HIV-1 BF1 sequences (BRMA94, BRMA95, BRPI96, BRMA242, BRGO4122, and BRGOMI744) that formed Cluster #1 (described in Reis et al., 2017) we were able to obtain the near-full-length sequence of the BRGOMI744 strain (411-9617 bp) from Goiás State. For the BRMA94 strain, from Maranhão State, only partial genome sequence of fragments were obtained (1028–5930 bp, 6004–7765 bp, relative to HXB2 genome positions). The bootscanning analysis of the near-full genome sequence of the BF1 strain BRGOMI744 showed a complex pattern comprising six recombination breakpoints which divide the genome into seven subregions (**Figure 8**). The following subtypes and subregions were identified in this genome: subtype B (626–910 bp; 2558–3693 bp; 5808–6384 bp; and 7317–9493 bp) and subtype F1 (911–2557 bp; 3694–5807 bp; and 6385–7316 bp) all of them relative to HXB2 genome positions. The partial genome fragments of the BRMA94 strain presented the same mosaic structure as described in the BRGOMI744 strain.

The Cluster #3 (described in Reis et al., 2017) comprised four isolates (BRGO3153, BRGO4057, BRMT1319, and BRMT2835). The full-genome sequence of the BRMT1319 strain showed that this URF_BF1 is mainly composed of subtype B. The bootscanning analysis of this genome showed six recombination breakpoints, dividing the genome into seven subregions (**Figure 9**). The subtypes and subregions identified in this genome were the following: subtype B (626–1218 bp; 1947–2307 bp; 2525–2697 bp; and 2876–9423 bp) and subtype F1 (1219–1946 bp; 2308–2524 bp; and 2698–2875 bp) all of them relative to HXB2 genome positions.

DISCUSSION

Our study with near-full and full-length genome sequences of recombinant viruses describes two new HIV-1 CRFs: the BF1C CRF81_cpx and the CRF99_BF1 that were identified among over 800 patients living in six Brazilian states. These states

are located across three different geographic areas where the prevalent subtype B cocirculates with minor subtypes F1 and C. Our study also characterized two URFs BF1 in this geographic area. This is the first description of a CRF_cpx in Brazil and it is the first among 23 CRF_cpx already reported at Los Alamos HIV Sequence Database, that does not contain subtype A (http://www.hiv.lanl.gov, last accessed December 2018). So far, all CRF_cpx described worldwide presented a partial subtype A genome, either "pure" or as CRF01 (subtypes A, E) or CRF02 (subtypes A, G) (http://www.hiv.lanl.gov, last accessed December 2018). The complex CRF characterized here contains nine breakpoints and 10 subregions of subtypes B, F1, and C. Our data suggest that the CRF81_cpx originated almost 20 years ago in 1999 and that it circulates in a wide geographic area including the Central Western region, where it probably originated, and the Southern Brazilian (Paraná state) region.

In addition, our analyses provided evidence that this newly described BF1C CRF81_cpx also circulates in Italy, where the HIV/AIDS epidemic has been characterized by an increasing genetic diversity of mainly non-B subtype strains (Lai et al., 2010). A previous phylogeographic study indicated an association between the Italian and the South American HIV/AIDS epidemic suggesting that the main source of the Italian subtype C epidemic is associated with interactions between Italian heterosexual and South American homosexual males (Lai et al., 2014). In the Italian study, although patients infected with subtype C mainly referred themselves as heterosexuals, homo- or bisexual contacts were considered the source of their infection (Lai et al., 2014). Although the investigation of the sexual links involved in the origin of the CRF81_cpx was out of the scope of this study, one of patients harboring the BF1C CRF81_cpx reported himself as MSM, suggesting a possible role of this category in the origin of this mosaic CRF.

Corroborating a possible link between the Italian and the South American HIV/AIDS epidemics, in our dataset we have found a CRF60_BC-like cluster represented by strains circulating in Northeastern and Central Western Brazil. The CRF60_BC was originally identified in Southern Italy after an outbreak among 22 patients, mostly highly educated, young MSM (Monno et al., 2012; Simonetti et al., 2014). Phylogenetic analysis of full-length genome of CRF60_BC revealed the South American origin of the C subtype parental strain (Simonetti et al., 2014). Similarly to the newly described CRF81_cpx, that appears to circulate in Brazil and Italy, our findings of CRF60_BF-like sequences support the connection between the HIV/AIDS epidemic in these two countries. It has been proposed that the presence of about 400,000 South American immigrants in Italy and of a significant and unstable population of transgender sex workers (around 10,000 individuals, 60% of them from South America) strenghthen this genetic association (Lai et al., 2014). Further studies are necessary to fully evaluate the spread and the geographic area of circulation of the CRF60_BC in Brazil.

Whereas we have found evidence of the spread of CRF60_BC in Brazil, up to now only one CRF containing subtypes B and C, the CRF31_BC was described in Brazil, in the Southern region, where subtype C is a main genetic variant (Santos et al., 2006). In our study we have found a CRF31_BC-like cluster indicating the dissemination of this CRF or a variant derived from it, from Southern to Central Western and Northeastern Brazil. In this study we also characterized the new CRF99_BF1 and two new URFs BF1. Currently there are 98 HIV-1 CRFs reported globally (http://www.hiv.lanl.gov, last accessed December 2018). So far, among the 15 CRF_BF reported at the Los Alamos HIV Sequence Database, the majority was described in different geographic regions of Brazil: CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF, CRF46_BF, CRF70_BF, CRF71_BF, CRF72_BF, and the CRF90_BF1 which was recently described by our group in the Central Western region (Sanabani et al., 2006, 2010; Guimarães et al., 2008; Pessôa et al., 2014a,b; Reis et al., 2017). Several URFs BF1 and URFs BC have been described by full and near-full-length genome sequencing studies in Brazil and this approach is necessary to define the real contribution of these mosaic forms in the Brazilian epidemic (Barreto et al., 2006; Sanabani et al., 2006, 2010; Passaes et al., 2009; Pessôa et al., 2016; Marques et al., 2018).

CONCLUSION

The description of the first CRF_cpx composed of subtypes B, F1, and C identified in Brazil shows that more complex HIV-1 recombinant variants are circulating in the country and also in Italy. Additionally the description of the new CRF99_BF1 and of two new URFs BF1 confirms the wide generation and spread of BF1 intersubtype recombinants in the Brazilian HIV/AIDS epidemic. Our results reinforce the need to expand HIV-1 full-length genomic studies in Brazil and worldwide in order to estimate the overall proportion of intersubtype recombinants in the HIV/AIDS epidemic.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Resolução 196/96 da CONEP/MS, Comite de Ética e pesquisa do Hospital de Doenças Tropicais/GO; CEPMHA/HC/UFG no186; 073/05, Comite de Ética em Pesquisa Médica Humana e Animal; No435/CEP-HUJM/07, Comite de Ética em Pesquisa do Hospital Universitário Júlio Miller; CEP-UESPI 022/2011, Comite de Ética em pesquisa da Universidade Estadual do Piauî with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Comite de Ética e pesquisa do Hospital de Doenças Tropicais, Comite de Ética em Pesquisa Médica Humana e Animal, Comite de Ética em Pesquisa do Hospital Universitário Júlio Miller, Comite de Ética em pesquisa da Universidade Estadual do Piauî.

AUTHOR CONTRIBUTIONS

MS conceived, designed and received funding for the study. MS and MR collected and analyzed the epidemiological data.

MR, MG and GB performed the phylogenetic analyses. GB performed the evolutionary analyses. MR performed all figures and tables. MS and MR wrote the manuscript. All authors discussed and reviewed the manuscript.

FUNDING

This work was supported by the "Programa de Apoio a Núcleos de Excelência/PRONEX; Fundação de Amparo à Pesquisa do Estado de Goiás/FAPEG; Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq (Grant Number: 201210267000801 to MS) and by the "Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq Universal (Grant Number: 481208/2012-7 to MS). MR was supported by a scholarship from FAPEG (Grant Number: 201410267000598). MS and MG are recipients of fellowships from CNPq (Grant Number 308381/2015-7 to MS and Grant Number 307972/20014-3 to MG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00097/full#supplementary-material

REFERENCES

Alcântara, K. C., Lins, J. B., Albuquerque, M., Aires, L. M., Cardoso, L. P., Minuzzi, A. L., et al. (2012). HIV-1 mother-to-child transmission and drug resistance among Brazilian pregnant with high access to diagnosis and prophylactic measures. *J. Clin. Virol.* 54, 15–20. doi: 10.1016/j.jcv.2012.01.011

Barreto, C. C., Nishyia, A., Araujo, L. V., Ferreira, J. E., Busch, M. P., Sabino, E. C., et al. (2006). Trends in antiretroviral drug resistance and clade distributions among HIV-1infected blood donors in Sao Paulo. Brazil. *J. Acquir. Immune. Defic. Syndr.* 41, 338–341. doi: 10.1097/01.qai.0000199097.88344.50

Cardoso, L. P., da Silveira, A. A., Francisco, R. B., da Guarda Reis, M. N., and Stefani, M. M. (2011). Molecular characteristics of HIV type 1 infection among prisoners from central western Brazil. *AIDS Res. Hum. Retroviruses* 27, 1349–1353. doi: 10.1089/aid.2011.0153

Cardoso, L. P., Pereira, G. A., Viegas, A. A., Schmaltz, L. E., and Stefani, M. M. (2010). HIV-1 primary and secondary antiretroviral drug resistance and genetic diversity among pregnant women from central Brazil. *J. Med. Virol.* 82, 351–357. doi: 10.1002/jmv.21722

Cardoso, L. P., Queiroz, B. B., and Stefani, M. M. (2009). HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naïve patients from Central West Brazil. *J. Clin. Virol.* 46, 134–139. doi: 10.1016/j.jcv.2009.07.009

Cardoso, L. P. V., and Stefani, M. M. A. (2009). High level of multidrug resistance mutations in HIV type 1 pol gene and resistance-associated mutations to enfuvirtide (T-20) among antiretroviral-experienced patients from central Brazil. *AIDS Res. Hum. Retroviruses* 25, 943–950. doi: 10.1089/aid.2009.0060

Carvalho, B. C., Cardoso, L. P., Damasceno, S., and Stefani, M. M. (2011). Moderate prevalence of transmitted drug resistance and interiorization of HIV type 1 subtype C in the inland North State of Tocantins. Brazil. *AIDS Res. Hum. Retroviruses* 27, 1081–1087. doi: 10.1089/AID.2010.0334

Cavalcanti, A. M., Brito, A. M., Salustiano, D. M., Lima, K. O., Silva, S. P., and Lacerda, H. R. (2012). Recent HIV infection rates among HIV positive patients seeking voluntary counseling and testing centers in the metropolitan region of recife - PE. Brazil. *Braz. J. Infect. Dis.* 16, 157–163. doi: 10.1590/S1413-86702012000200009

da Costa, Z. B., de Lima, Y. A., Martelli, C. M., and Stefani, M. M. (2013). Transmitted HIV resistance among pregnant young women infected with HIV-1 in Brazil. *AIDS Patient Care STDS* 27, 439–441. doi: 10.1089/apc.2012.0448

da Silveira, A. A., Cardoso, L. P., Francisco, R. B., and de Araújo Stefani, M. M. (2012). HIV type 1 molecular epidemiology in pol and gp41 genes among naïve patients from mato grosso do sul state, central western Brazil. *AIDS Res. Hum. Retroviruses* 28, 304–307. doi: 10.1089/aid.2011.0128

Delatorre, E., Couto-Fernandez, J. C., and Bello, G. (2017). HIV-1 genetic diversity in northeastern Brazil: high prevalence of non-B subtypes. *AIDS Res. Hum. Retroviruses* 33, 639–647. doi: 10.1089/AID.2017.0045

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

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307–1320.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Faria, N. R., Rambaut, A., Suchard, A. A., Baele, G., Bedford, T., Ward, M. J., et al. (2014). HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346, 56–61. doi: 10.1126/science.1256739

Ferreira, A. S., Cardoso, L. P., and Stefani, M. M. (2011). Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso state, central western Brazil. *J. Med. Virol.* 83, 1301–1307. doi: 10.1002/jmv.22128

Gaspareto, K. V., Mello, F. M., Dias, J. R., Meneguetti, V. A., Storti, M. E., Ferreira, J. L., et al. (2012). Genetic diversity and primary resistance among HIV-1-positive patients from maringá, paraná, Brazil. *Rev. Inst. Med. Trop. Sao Paulo* 54, 207–213. doi: 10.1590/S0036-46652012000400005

Gräf, T., Machado Fritsch, H., de Medeiros, R. M., Maletich Junqueira, D., Esteves de Matos Almeida, S., and Pinto, A. R. (2016). Comprehensive characterization of HIV-1 molecular epidemiology and demographic history in the BRAZILian region most heavily affected by AIDS. *J. Virol.* 26, 8160–8168. doi: 10.1128/JVI.00363-16

Gräf, T., and Pinto, A. R. (2013). The increasing prevalence of HIV 1 subtype C in southern Brazil and its dispersion through the continent. *Virology* 435, 170–178. doi: 10.1016/j.virol.2012.08.048

Guimarães, M. L., Couto-Fernandez, J. C., Eyer-Silva, W. A., Teixeira, S. L., Chequer-Fernandez, S. L., Morgado, M. G., et al. (2010). Analysis of HIV-1 BF pr/rt recombinant strains from rio de janeiro/Brazil reveals multipleunrelated mosaic structures. *Infect. Genet. Evol.* 10, 1094–1100. doi: 10.1016/j.meegid.2010.07.001

Guimarães, M. L., Eyer-Silva, W. A., Couto-Fernandez, J. C., and Morgado, M. G. (2008). Identification of two new CRF_BF in rio de janeiro state, Brazil. *AIDS* 22, 433–435. doi: 10.1097/QAD.0b013e3282f47ad0

Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D., and Dougherty, J. P. (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* 74, 1234–1240. doi: 10.1128/JVI.74.3.1234-1240.2000

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Lai, A., Ciccozzi, M., Franzetti, M., Simonetti, F. R., Bozzi, G., Binda, F., et al. (2014). Local and global spatio-temporal dynamics of HIV-1 subtype F1. *J. Med. Virol.* 86, 186–192. doi: 10.1002/jmv.23783

Lai, A., Riva, C., Marconi, A., Balestrieri, M., Razzolini, F., Meini, G., et al. (2010). Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. *HIV Med.* 11, 593–602. doi: 10.1111/j.1468-1293.2010.00832.x

Lima, K., de Souza Leal, É, Cavalcanti, A. M., Salustiano, D. M., de Medeiros, L. B., da Silva, S. P., et al. (2016a). Epidemiological, clinical and antiretroviral susceptibility characterization of human immunodeficiency virus subtypes B

and non-B in pernambuco, northeast Brazil. *PLoS One* 11:e0155854. doi: 10.1371/journal.pone.0158192

Lima, Y. A., Cardoso, L. P., Reis, M. N., and Stefani, M. M. (2016b). Incident and long term HIV-1 infection among pregnant women in Brazil: transmitted drug resistance and mother-to-child transmission. *J. Med. Virol.* 88, 1936–1943. doi: 10.1002/jmv.24540

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Marques, B. C. L., Morgado, M. G., and Guimarães, M. L. (2018). Potential overestimation of HIV-1 sub-subtype F1 circulation in Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* 113:e170483. doi: 10.1590/0074-02760170483

Monno, L., Brindicci, G., Lai, A., Punzi, G., Altamura, M., Simonetti, F. R., et al. (2012). An outbreak of HIV-1 BC recombinants in Southern Italy. *J. Clin. Virol.* 55, 370–373. doi: 10.1016/j.jcv.2012.08.014

Moura, M. E., da Guarda Reis, M. N., Lima, Y. A., Eulálio, K. D., Cardoso, L. P., and Stefani, M. M. (2015a). HIV-1 transmitted drug resistance and genetic diversity among patients from Piauí State. Northeast Brazil. *J. Med. Virol.* 87, 798–806. doi: 10.1002/jmv.24087

Moura, M. E., da Guarda Reis, M. N., Lima, Y. A., Eulálio, K. D., Cardoso, L. P., and Stefani, M. M. (2015b). Low rate of transmitted drug resistance may indicate low access to antiretroviral treatment in Maranhão State, northeast Brazil. *AIDS Res. Hum. Retroviruses* 31, 250–254. doi: 10.1089/aid.2014.0261

Nei, M., and Kumar, S. (2002). *Molecular Phylogenetics and Evolution*. Oxford: Oxford University.

Passaes, C. P., Guimarães, M. L., Bello, G., and Morgado, M. G. (2009). Near full-length genome characterization of HIV type 1 unique BC recombinant forms from Southern Brazil. *AIDS Res. Hum. Retroviruses* 25, 1339–1344. doi: 10.1089/aid.2009.0167

Pessôa, R., Carneiro Proietti, A. B., Busch, M. P., and Sanabani, S. S. (2014a). Identification of a novel HIV-1 circulating recombinant form (CRF72_BF1) in deep sequencing data from blood donors in Southeastern Brazil. *Genome Announc.* 2, e386–e414. doi: 10.1128/genomeA.00386-14

Pessôa, R., Watanabe, J. T., Calabria, P., Felix, A. C., Loureiro, P., Sabino, E. C., et al. (2014b). Deep sequencing of HIV-1 near full-length proviral genomes identifies high rates of BF1 recombinants including two novel circulating recombinant forms (CRF) 70_BF1 and a disseminating 71_BF1 among blood donors in Pernambuco, Brazil. *PLoS One* 17:e112674. doi: 10.1371/journal.pone.0112674

Pessôa, R., Loureiro, P., Esther Lopes, M., Carneiro-Proietti, A. B., Sabino, E. C., Busch, M. P., et al. (2016). Ultra-deep sequencing of HIV-1 near full-length and partial poviral genomes reveals high genetic diversity among BRAZILian bood donors. *PLoS One* 11:e0152499. doi: 10.1371/journal.pone.0152499

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Rambaut, A., and Drummond, A. (2007). *Tracer v1.6*. Available at: http://tree.bio.ed.ac.uk/software/tracer/

Reis, M. N. D. G., Bello, G., Guimarães, M. L., and Stefani, M. M. A. (2017). Characterization of HIV-1 CRF90_BF1 and putative novel CRFs_BF1 in Central West, North and Northeast BRAZILian regions. *PLoS One* 12:e0178578. doi: 10.1371/journal.pone.0178578

Sanabani, S., Kleine Neto, W., Kalmar, E. M., Diaz, R. S., Janini, L. M., Sabino, E. C., et al. (2006). Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. *Infect. Genet. Evol.* 6, 368–377. doi: 10.1016/j.meegid.2006.01.003

Sanabani, S. S., Pastena, E. R., Neto, W. K., Martinez, V. P., and Sabino, E. C. (2010). Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. *Virol. J.* 7:74. doi: 10.1186/1743-422X-7-74

Santos, A. F., Sousa, T. M., Soares, E. A., Sanabani, S., Martinez, A. M., Sabino, E. C., et al. (2006). Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* 20, 2011–2019.

Silveira, J., Santos, A. F., Martínez, A. M., Góes, L. R., Mendoza-Sassi, R., Muniz, C. P., et al. (2012). Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J. Clin. Virol.* 54, 36–41. doi: 10.1016/j.jcv.2012.01.017

Simonetti, F. R., Lai, A., Monno, L., Binda, F., Brindicci, G., Punzi, G., et al. (2014). Identification of a new HIV-1 BC circulating recombinant form (CRF60_BC) in Italian young men having sex with men. *Infect. Genet. Evol.* 23, 176–181. doi: 10.1016/j.meegid.2014.02.007

Soares, E. A., Martínez, A. M., Souza, T. M., Santos, A. F., Da Hora, V., Silveira, J., et al. (2005). HIV-1 subtype C dissemination in southern Brazil. *AIDS* 19(Suppl. 4), S81–S86. doi: 10.1097/01.aids.0000191497.00928.e4

Suchard, M. A., and Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25, 1370–1376. doi: 10.1093/bioinformatics/btp244

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thompson, J. D., Higgins, D. S., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Worobey, M., and Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* 80, 2535–2543. doi: 10.1099/0022-1317-80-10-2535

Xia, X., and Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* 92, 371–373. doi: 10.1093/jhered/92.4.371

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 February 2019
doi: 10.3389/fmicb.2019.00287

Exploring Evolutionary and Transmission Dynamics of HIV Epidemic in Serbia: Bridging Socio-Demographic With Phylogenetic Approach

Luka Jovanović[1†], Marina Šiljić[1†], Valentina Ćirković[1], Dubravka Salemović[2], Ivana Pešić-Pavlović[3], Marija Todorović[1], Jovan Ranin[2], Djordje Jevtović[2] and Maja Stanojević[1]*

[1] Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia, [2] Infectious and Tropical Diseases University Hospital, Clinical Centre of Serbia, Belgrade, Serbia, [3] Virology Laboratory, Microbiology Department, Clinical Centre of Serbia, Belgrade, Serbia

OPEN ACCESS

Edited by:
Joris Hemelaar,
University of Oxford, United Kingdom

Reviewed by:
Dimitrios Paraskevis,
National and Kapodistrian University of Athens, Greece
Guido van Marle,
University of Calgary, Canada

***Correspondence:**
Maja Stanojević
maja.stanojevic@med.bg.ac.rs

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 14 October 2018
Accepted: 04 February 2019
Published: 25 February 2019

Citation:
Jovanović L, Šiljić M, Ćirković V, Salemović D, Pešić-Pavlović I, Todorović M, Ranin J, Jevtović D and Stanojević M (2019) Exploring Evolutionary and Transmission Dynamics of HIV Epidemic in Serbia: Bridging Socio-Demographic With Phylogenetic Approach. Front. Microbiol. 10:287. doi: 10.3389/fmicb.2019.00287

Previous molecular studies of Serbian HIV epidemic identified the dominance of subtype B and presence of clusters related HIV-1 transmission, in particular among men who have sex with men (MSM). In order to get a deeper understanding of the complexities of HIV sub-epidemics in Serbia, epidemic trends, temporal origin and phylodynamic characteristics in general population and subpopulations were analyzed by means of mathematical modeling, phylogenetic analysis and latent class analysis (LCA). Fitting of the logistic curve of trends for a cumulative annual number of new HIV cases in 1984–2016, in general population and MSM transmission group, was performed. Both datasets fitted the logistic growth model, showing the early exponential phase of the growth curve. According to the suggested model, in the year 2030, the number of newly diagnosed HIV cases in Serbia will continue to grow, in particular in the MSM transmission group. Further, a detailed phylogenetic analysis was performed on 385 sequences from the period 1997–2015. Identification of transmission clusters, estimation of population growth (Ne), of the effective reproductive number (Re) and time of the most recent common ancestor (tMRCA) were estimated employing Bayesian and maximum likelihood methods. A substantial proportion of 53% of subtype B sequences was found within transmission clusters/network. Phylodynamic analysis revealed Re over one during the whole period investigated, with the steepest slopes and a recent tMRCA for MSM transmission group subtype B clades, in line with a growing trend in the number of transmissions in years approaching the end of the study period. Contrary, heterosexual clades in both studied subtypes – B and C – showed modest growth and stagnation. LCA analysis identified five latent classes, with transmission clusters dominantly present in 2/5 classes, linked to MSM transmission living in the capital city and with the high prevalence of co-infection with HBV and/or other STIs.

Presented findings imply that HIV epidemic in Serbia is still in the exponential growth phase, in particular, related to the MSM transmission, with estimated steep growth curve until 2030. The obtained results imply that an average new HIV patient in Serbia is a young man with concomitant sexually transmitted infection.

Keywords: HIV-1, subtype B, MSM, transmission clusters, model, phylodynamics, latent class analysis, Serbia

INTRODUCTION

The pandemic of human immunodeficiency virus (HIV) infection has developed in numerous heterogeneous sub-epidemics worldwide, substantially influenced by patterns of human migration and globalization. Similar to what happened worldwide, Serbian HIV epidemic has evolved through several phases. Serbia is a low HIV-1 prevalence country. According to the national surveillance data, since the first HIV cases emerged in 1985 until the end of 2017, over 3,500 people in Serbia were living with HIV/AIDS, without substantial reduction in the yearly number of newly diagnosed cases[1]. In the beginning, the HIV epidemic in Serbia spread mostly among people who inject drugs (PWID), until the mid-1990s, when the heterosexual transmission route became more prevalent. In recent years another shift in the dominant transmission route was observed, leading to the current situation where HIV epidemic is driven by transmission among men having sex with men (MSM), comprising the majority of new HIV diagnoses, with the yet ongoing presence of heterosexual transmission.

Previous molecular studies of Serbian HIV epidemic identified subtype B as the dominant HIV-1 type, present in over 90% of individuals diagnosed between 1997 and 2011 (Stanojevic et al., 2002, 2014; Siljic et al., 2013; Magiorkinis et al., 2016). Clusters related HIV-1 transmission was observed, in particular among MSM transmission group and linked to high prevalence of concomitant sexually transmitted infections (STIs) (Siljic et al., 2013). This rapid expansion of HIV-1 infection among MSM transmission group and high prevalence of reported STIs underlined the need for detailed molecular investigation of the network structure and dynamics of viral transmission in this and other groups.

In order to get a deeper understanding of the complexities of HIV sub-epidemics in Serbia, epidemic trends, temporal origin and phylodynamic characteristics in general population and subpopulations were analyzed by means of mathematical modeling, phylogenetic analysis, and latent class analysis (LCA).

MATERIALS AND METHODS

Logistic Growth Modeling

Logistic growth modeling was performed based on data about new HIV cases, as well as the basic related demographic data in the period 1984–2016, available from the Institute of Public Health of Serbia "Dr. Milan Jovanovic Batut" and the annual HIV reports of the European Centre for Disease Control (see footnote 1; ECDC, 2017).

Fitting of the logistic curve of trends for a cumulative annual number of new HIV cases in the period 1984–2016, in general population, heterosexual transmission group, PWID transmission group and MSM transmission group, was performed with the software NLREG, v 6.6[2]. The logistic regression model assumes the existence of a maximum population size which a given environment can accommodate (carrying capacity alias K) and a biphasic growth rate creating two stages of growth: (i) an early exponential phase and (ii) a late phase of the plateau (Vandermeer, 2010). In the context of an HIV epidemic the exponential phase of logistic growth could illustrate the absence of large scale preventive measures or the presence of inefficient preventive policy (Fonseca and Bastos, 2007). The following logistic growth model was used: $y = K/(1 + \exp^*(a + b^*x))$, where K represents carrying capacity, a and b represent parameters that shape and scale the function (Sherrod, 2010). NLREG provides an estimation of parameters K, a and b with the strongest statistical back-up. It also provides several diagnostic statistical tests and variables which describe the statistical strength of the proposed model such as (i) probability t (prob. t) is a probability that the estimated parameter (a, b and K) is 0; prob. t less than 0.05 is considered as a good estimation of parameter significance for the model; (ii) probability f (prob. f) is the probability that all of the regression parameters are 0, and it ranges from 0 to 1; prob. f less than 0.05 illustrates an overall statistical significance of the model. The model obtained was used to predict further trends of HIV epidemic in general population and by transmission groups (MSM, PWID, heterosexual transmission) for the period 2017–2030. This analysis provided a general overview of the epidemiological background for further transmission clusters analysis.

Phylogenetic Analyses

Study Population and Sequence Dataset

The earliest viral sequences from Serbian epidemic date from 1997. For the time period considered, 1997 to 2011, data and sequences were gathered from previous molecular studies of HIV epidemic in Serbia. Complete dataset of these sequences is available at the NCBI database, accession numbers are given in **Supplementary File 1**). Additionally, this study included blood samples collected from 2012 to 2015, from consenting HIV seropositive adults, both therapy naïve and therapy experienced, followed at the Centre for HIV/AIDS, University Hospital for Infectious and Tropical Diseases in Belgrade. Blood samples from

[1]http://www.batut.org.rs/index.php?category_id=17

[2]http://www.nlreg.com

HIV-infected patients were sent for drug resistance testing as part of patients' routine follow-up. The study was approved by the Ethical Committee of the University of Belgrade Faculty of Medicine 29/V-11.

In brief, HIV-1 genomic RNA was extracted from 140 μL of stored plasma specimens using the QIAmp Viral RNA Mini kit (Qiagen, Hilden, Germany). Reverse transcription and nested PCR amplification of partial pol gene were performed using One Step RNA PCR Kit (Qiagen, Hilden, Germany), Thermo scientific dreamTaq PCR master mix (2X) (Applied Biosystems, Foster City, CA, United States) and HIV-1 specific primer sets (Snoeck et al., 2005). Upon successful amplification and purification with MinElute Purification Kit (Qiagen, Hilden, Germany), according to the manufacturer's instruction, PCR products were subjected to direct sequencing of both the sense (forward) and antisense strands (reverse) by BigDye®Terminator v3.1 (Applied Biosystems, Foster City, CA, United States).

Obtained sequences were visually inspected, manually edited and then assembled with SeqScape HIV-1 Genotyping System Software v 2.5 (Applied Biosystems, Foster City, CA, United States). Sequences generated in this research were deposited in the NCBI database (accession numbers are shown in **Supplementary File 1**).

Reference sequences of different subtypes, known to be present locally in Serbia and in the Balkan region, encompassing subtypes B, C, G, A, F, as well as circulating recombinant forms (CRFs), CRF01_AE and CRF02-AG, were downloaded from the Los Alamos database[3]. Additionally, NCBI BLAST search was performed for each sequence found to belong to a transmission cluster (according to criteria described in detail further). For each query sequence, five sequences were included in the tree reconstruction based on the highest similarity score as obtained by NCBI BLAST search. In total, 175 sequences were included in the analyses after BLAST search. Furthermore, a group of control/background sequences was also included, in the context of the origin and spread of HIV from the isolates of North America, West Europe, and Balkan, with the clear defined subtype and the time of sampling available at the NCBI database (accession numbers are shown in **Supplementary File 1**).

Data on clinical characteristics (CDC stage, baseline CD4 count, coinfections and other sexually transmitted illnesses), epidemiological background (age, gender, date of diagnosis, residence and level of education) and risk for acquiring HIV infection (most probable HIV transmission route) were used as a covariate of transmission analysis.

HIV Subtyping

Subtyping of all sequences included in the study was firstly performed using the REGA HIV-1 subtyping tool version 3 (REGA v3)[4] and the Los Alamos HIV database (see footnote 3). The Rega subtyping tool is based on phylogenetic analysis in order to take into account the epidemiological and evolutionary relationships among subtypes (Robertson et al., 2000). Subtyping of all pol gene sequences included in this research was further performed by construction of ML and NJ phylogenetic tree, under appropriate models, using the PAUP and MEGA software, with reference sequences of different subtypes, downloaded from HIV-1 Los Alamos Database (LANL[5]). Sequences that were not unambiguously classified by REGA subtyping tool or that were assigned to a different subtype from the one specified in the Los Alamos database were removed from the data set.

Phylogenetic Trees Reconstruction

To eliminate the influence of antiretroviral drug selective pressure on viral evolution, we made a codon-stripped sequence alignment by removing drug resistance-associated codons identified by the Stanford Drug Resistance Database[6] (Wensing et al., 2017). Analysis was also performed without codons removing. Both analyses gave congruent results with no evidence of the selection pressure of antiretroviral drugs to the clustering patterns. Sequences were aligned using the Clustal W algorithm implemented in MEGA v 6[7] and were manually edited.

The Bayesian phylogenetic tree was reconstructed through Bayesian inference of phylogeny by using MrBayes software (Ronquist and Huelsenbeck, 2003). A Markov Chain Monte Carlo (MCMC) search was made for 10 × 106 generations using tree sampling every 100th generation and a burn-in fraction of 20%. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and a posterior consensus tree was generated after a 25% burn-in.

Maximum likelihood (ML) tree reconstruction using PAUP was performed under GTR model with discrete gamma rates heterogeneity among sites and the proportion of invariable sites, as selected by jModelTest (Swofford, 1999; Guindon and Gascuel, 2003; Posada, 2009).

Phylogenetic analyses were conducted: (i) to confirm subtype assignment performed with REGA subtyping tool (ii) to identify the presence of local transmission networks (phylogenetic clusters). and (iii) to perform in-depth reconstruction of demographic and transmission history of the HIV epidemic in Serbia.

Analysis of Transmission Clusters

A set of analyses was performed for a definition of HIV transmission clusters, following the strategy of "statistical support of clades plus similarity" that proved to be suitable in our previous investigation of transmission chains in Serbia (Siljic et al., 2013, 2017). Evidence of transmission cluster was characterized by two criteria sets: according to the first set of criteria, transmission clusters were assigned as those monophyletic phylogenetic clades consisting of three or more sequences, fulfilling the conditions of genetic distance of 1.5% or less, with minimal bootstrap support of 90%, and the Bayesian posterior probability of higher than 0.9. Genetic distances were first determined using PAUP* version 4.0 software, based on ML analysis and under general time reversible model (GTR) with gamma distribution and proportion of invariable sites

[3]https://www.hiv.lanl.gov/content/index

[4]http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/

[5]www.hiv.lanl.gov

[6]https://hivdb.stanford.edu/hivdb/by-mutations/

[7]https://www.megasoftware.net/

(Swofford, 1999; Guindon and Gascuel, 2003), selected by jModelTest statistical analyses (Posada, 2009). Branch supports were estimated using bootstrap analysis with 1,000 replicates. Phylogenetic trees were visualized in FigTree program version 1.3.1[8]. Second criteria sets, including those clades with bootstrap support over 75% and with genetic distance of less than 4%, were further analyzed.

For each sequence found within a cluster according to the second criteria set, 5 most similar sequences were identified using BLAST analysis[9] and included in tree reconstruction – subsistence of the initial local clusters after inclusion of BLAST identified sequences in the tree reconstruction led to their designation as true transmission clusters.

Bayesian Molecular-Evolution Analyses

Timed evolutionary histories, demographic history, and reproductive effective number estimates

Subtype B MSM transmission network and the two most expanded MSM transmission clusters (composed of 15 and 11 sequences) identified in the previous analyses, were subjected to detailed temporal and phylodynamic reconstruction aimed to infer demographic and evolutionary histories of these monophyletic groups.

Additionally, two heterosexual monophyletic clades were analyzed, one each of subtypes B and C, in order to analyze these sub-epidemics. The studied subtype C clade was supported by 100% value of bootstrap support and composed of sequences from heterosexuals only, whereas for subtype B, phylodynamic analyses were performed on a monophyletic clade, with lower than predefined bootstrap value, but composed of sequences dominantly isolated from heterosexuals. Moreover, as a method of validation against the possible bias due to low bootstrap support seen in subtype B heterosexual clade, an additional monophyletic clade with similar bootstrap support, comprised of sequences from MSM transmission group, was chosen to be analyzed. The results of these analyses and detailed methods are presented in the **Supplementary File 2**.

Phylodynamic analyses in this research were performed as follows: estimations of evolutionary and demographic parameters were performed in BEAST software package v 1.7.5 and encompassed selection for site model, demographic model, and clock model prior to run MCMC chains. In the first step, the Bayesian skyline plot method (Drummond et al., 2005, 2012), was used to estimate effective population size (Ne), and to run a demographic analysis in Tracer v1.6[10]. In the second step, as for the coalescent priors, two different demographic models were compared: exponential and logistic growth; estimates of the population growth rate were then obtained under the model that provided the best fit to the demographic signal in each data set. The best fit coalescent model was chosen by means of a Bayes factor (BF), using marginal likelihoods, determined by Tracer version 1.6. Each analysis was performed under GTR+G+I, as the best fit nucleotide substitution model, and uncorrelated lognormal relaxed clock, as shown to be the most suitable for HIV datasets. MCMC chains were run for 5 × 107 generations for each data set, with a burn-in of 10%. BEAST output was analyzed using TRACER v1.6, with uncertainty in parameter estimates reflected as the 95% highest probability density (HPD). Convergence of parameters was assessed through the ESS after excluding an initial 10% for each run. All parameter estimates for each run showed ESS values >200. A graphical representation of the effective number of infections through time was generated by using TRACER v1.6.

In order to estimate the time of the most recent common ancestors for the selected clades, molecular clock analysis was performed using a Bayesian MCMC coalescent method, as implemented in BEAST v1.8.1 (Drummond et al., 2003; Drummond and Rambaut, 2007). We used GTR nucleotide substitution model with six category gamma distributed rate variation among sites and two partitions in the codon positions (the best fitting model for all three datasets according to jModelTest). Bayes factor analysis, performed in Tracer v.1.5, showed that the lognormal relaxed clock with the Bayesian Skyline was the best model, as indicated by a log Bayes Factor > 10 according to Tracer v1.5[11] – evidence of very strong statistical support (Suchard et al., 2001). Convergence of the Markov chain was assessed by program Tracer v 1.5[12] calculating the effective sample size (ESS) for each parameter.

The last part of phylodynamic analyses was the estimation of the effective reproductive number conducted in BEAST2 v 2.1.3 The analysis employed a general time-reversible substitution model with a gamma-distributed rate variation and proportion of invariant sites (GTR+Γ4+I), an uncorrelated lognormal relaxed molecular clock model (Drummond et al., 2006), and a Birth-Death Skyline Serial model (BDSKY) as a model for viral transmission (Stadler et al., 2012, 2013). The following prior distribution of the BDSKY model parameters was used: LogNorm (0; 10) for effective reproductive number R and LogNorm (0; 10) for the rate of becoming non-infectious δ. The BEAST2 analyses were run until all relevant parameters converged, with 20% of the MCMC chains discarded as burn-in. Statistical confidence was represented by values for the 95% HPD. To generate the log file, five independent MCMC runs of 2 × 108 chain length were combined with Log Combiner (Drummond et al., 2012). Sampling dates were used to infer the tree height and internal node ages in the maximum clade credibility (MCC) time-trees using BEAST2 (Bouckaert et al., 2014). Similarly to the log file, the time-trees from five independent runs of 2 × 108 chain length were combined with Log Combiner (Drummond et al., 2012) with 20% of the MCMC chains discarded as burn-in.

Latent Class Analysis

Latent class analysis is a statistical tool that explores underlying patterns of covariance in the data structure to identify subgroups or 'discrete classes' of participants' epidemiological, behavioral and clinical profiles. Class membership is inferred based on an individual's pattern of responses across a set of variables

[8]http://tree.bio.ed.ac.uk/software/figtree/

[9]http://blast.ncbi.nlm.nih.gov

[10]http://tree.bio.ed.ac.uk/software/tracer/

[11]http://beast.bio.ed.ac.uk/Tracer

[12]http://tree.bio.ed.ac.uk/software/tracer/

but not directly observed and therefore classes are considered to be latent. In this study, LCA was conducted to examine and identify participants risk profiles regarding 11 types of latent class indicators (categorical latent variable). Latent class indicators included: gender, transmission risk, age, residence, education level, coinfections with hepatitis B virus (HBV) and hepatitis C virus (HCV), presence of other STIs, CDC stage at the time of diagnosis, time-period of diagnosis and inclusion of sequence within clusters/network. These classification variables were selected to represent a range of established HIV risk factors in order to identify comprehensive HIV profiles, based on clinical, epidemiological, demographic and phylogenetic data. The covariates were separated into ordinal categories where each category was assigned a nominal value of 1 to 4.

Latent class analysis was performed in R software with polytomy variable LCA (poLCA) software package. PoLCA is a user-friendly package for the estimation of latent class models and latent class regression models in R available from the Comprehensive R Archive Network[13] (Linzer and Lewis, 2011; Schreiber, 2017). Based on the fact that this algorithm may locate a local, rather than global maximum, nrep was set to 10, which increased the probability that the global maximum log-likelihood would be located. Parameters used to select the optimal number of latent classes in LCA included the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), two most widely used parsimony measures (Akaike, 1973; Schwarz, 1978). We began with a 1-class model and increased the number of classes in each subsequent model seeking to minimize both the BIC and the AIC value before these values increased with the addition of another class.

Statistical Analysis

Results were analyzed by standard statistical tests. Categorical data were compared using the chi-square test and Fisher's exact test. Age and time of diagnosis of patients within network and clusters were statistically analyzed using chi-square test available at https://www.graphpad.com/quickcalcs/chisquared1.cfm.

RESULTS

Logistic Growth Modeling

In the period 1984–2016, the total cumulative number of new HIV cases in Serbia was 3590. The applied logistic growth model fitted accurately the growth trend in this time period (**Figure 1A**). Software estimated parameters *K*, *a*, and *b* equaling 4227 ± 271, 2.61 ± 0.07, -0.12 ± 0.008, respectively, with prob. t parameter 0.00001. Overall statistical significance was illustrated with prob. f = 0.00001.

Since data on transmission routes became available starting from 2003, the incidence of new HIV infections through MSM, PWID, and heterosexual transmission route in the period 2003–2016 was used in the analysis.

In this time period, a total of 2,785 new cases of HIV in MSM were detected. Again, the logistic growth model fitted the growth trend accurately (**Figure 1B**). The estimated function parameters *K*, *a*, and *b* were found to be 3377 ± 455, 3.12 ± 0.11, -0.15 ± 0.005, respectively, with prob. t = 0.00001. Overall statistical significance illustrated with prob. f 0.00001 showed strong statistical support for the proposed model. In the same time period, a total of 979 and 850 new cases of HIV in PWID and heterosexuals were found, respectively. The estimated function parameters *K*, *a* and *b* were 1018 ± 7, 1.74 ± 0.04, -0.8 ± 0.1 with prob. t = 0.00001 and with strong statistical significance of the model (prob. f = 0.00001) in PWID transmission group and 1576 ± 176, 1.06 ± 0.14, -0.08 ± 0.0059 with prob. t = 0.00001 and strong statistical support of the model (prob. f = 0.00001) in the heterosexual transmission group.

The fitted model was used to create trends of HIV epidemic for the period 2017–2030 in general population and by transmission groups. The modeled trend predicts approaching plateau in new HIV infections number by 2030 in general population (**Figure 2A**), whereas in MSM transmission group an exponential-like curve with no signs of a plateau in the studied time period was obtained (**Figure 2B**). Plateaued growth was observed in both PWID transmission group and heterosexual group (**Supplementary Figures 4**, **5**).

Phylogenetic Analyses

Phylogenetic analyses in the present study included a total of 385 local sequences, collected between 1997 and 2015 from both treatment naive and treatment experienced HIV-1 positive subjects. This pool represents around 11% of the total number of registered HIV+ individuals in Serbia in the same time period. Of these, the majority were male (81%), with a mean age of 37.5 years and reporting MSM contact (83.4%) as the most probable route of HIV acquisition. Epidemiological, demographic and clinical data are shown in **Table 1**.

In total, subtype B was confirmed to be the predominant one, accounting for 90.9% of cases (350/385), while the prevalence of non-B subtypes was 9.1% (35/385). Among non-B subtypes, subtype C was found with the highest prevalence, in 3.1% of samples (12/385), followed by subtype G in 2.1% (8/385) and subtypes A in 1.3 % (8/385). CRFs were detected in 2.6% (10/385) of the collected samples.

In phylogenetic trees reconstruction, both ML and Bayesian approach gave congruent results, revealing that 27.7% (107/385) of sequences were grouped within 19 transmission clusters, accomplishing predefined strict criteria (including nt distance of $\leq$1.5%), all within subtype B (**Figure 3**; labeled in dark blue). Moreover, 49 additional sequences were found grouped within a transmission network, a clade with high bootstrap support as well as posterior probability value but with a genetic distance of 8.2%, higher than the predefined cut off of 1.5% (**Figure 3**; labeled in red). This network contained sequences sampled in the time frame of 19 years and comprised a total of 63 sequences, encompassing thus 16.3% (63/385) of the total number of studied sequences. Taken together, the number of sequences within transmission clusters/network totaled to 205/385 (53%). Importantly, the majority of clusters, 15/19 (79%) were made solely of MSM transmission group sequences; with only 4/19 (21%) transmission clusters containing sequences of

[13] http://CRAN.R-project.org/package=poLCA

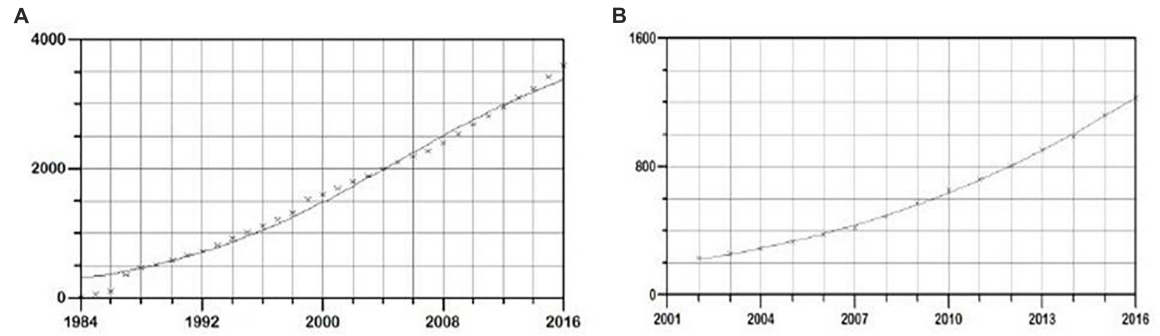


FIGURE 1 | Logistic growth model of new HIV infections in **(A)** general population and **(B)** MSM population in Serbia, in the period 1985–2016. On *y*-axis cumulative number of new HIV cases is shown, and on the *x* time period in years is shown.

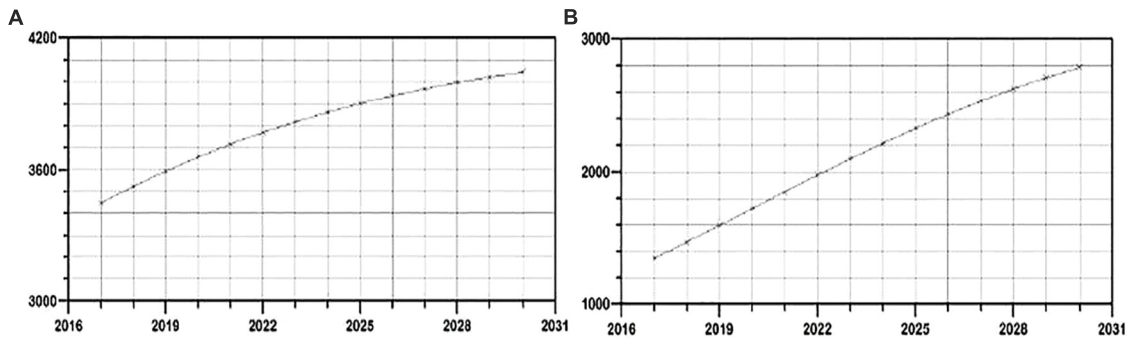


FIGURE 2 | Evaluation of the growth trends for the period 2017–2030 of new HIV infections in **(A)** general and **(B)** MSM population in Serbia.

heterosexuals. Sequences contributing to transmission clusters network were from individuals of a mean age of 35.3 years ($SD = 4.8$) and were mostly men ($N = 121/147$; 82.3%). Specifically, the large transmission network encompassed sequences from older patients of a mean age of 39.4 years, in contrast to MSM clusters with members mean age of 32.6 ($SD = 3.2$) with particular emphasis on the most expanded transmission cluster of 15 sequences with a mean age of 29.2 ($SD = 2.8$). Statistical evaluation using the chi-square test revealed that a significantly lower number of patients found in MSM clusters were diagnosed prior to 2006 ($p = 0.0184$) while a significantly higher number of them was younger than 35 ($p = 0.0388$). Among non-B clades, subtype C cluster contained sequences with the smallest nt distance (6.5%) and from heterosexual patients, with bootstrap support of 100; therefore we used this clade for further phylodynamic analysis. The observed subtype C genetic distance exceeded the pre-defined criteria for transmission clusters, hence it was not classified as one, however, downstream phylodynamic analyses were still performed, in order to be able to make comparisons to the relevant subtype B clades (**Figure 3**; labeled in green).

The time of the most recent common ancestor for transmission network (composed of 63 sequences) was dated to the beginning of nineties (1993 HPD: 1988–1995), while for the two MSM clusters of 15 and 11 sequences it was dated to 2008 (HPD: 2003–2013) and 2005 (HPD: 2000–2010), respectively. Dating analysis of two heterosexual clades suggested that the most recent common ancestor of the subtype B and subtype C clusters was present approximately in the 1990 (HPD: 1984–1996) and 1989 (HPD: 1985–1993), respectively.

A Bayesian Skyline plot and logistic growth analyses were performed for three MSM clades and two heterosexual clades, of subtype B and subtype C sequences. The highest exponential growth in almost 2 logs was identified for the most expanded transmission cluster, which was also the most recent one, in the whole analyzed period (**Figures 4A,B**). A demographic reconstruction of the transmission network and transmission cluster of 11 sequences showed slightly lower population growth, still with initial exponential growth in over one log which stabilized in the mid-decades of the 2000s, as reflected in a stationary phase approaching the present (**Figures 4C–F**). Bayesian skyline reconstruction and logistic growth analyses for the heterosexual clade of subtype B showed an initial increase of estimated effective population size from the beginning of the nineties until the late nineties, when the stationary phase started, and is still ongoing to this date (**Figures 5A,C**). Regarding subtype C clade, high exponential growth started in the mid-nineties and was followed by a stationary phase started at the beginning of the 2000s (**Figures 5B,D**).

Estimation of the effective reproductive number by birth-death skyline plot analyses showed significant differences among two sub-epidemics in Serbia, the MSM monophyletic clades

TABLE 1 | General epidemiological, demographic, and clinical data.

Gender	*N*	%
Male	312	81
Female	73	19
Transmission route		
MSM	302	78.4
Heterosexual	68	17.6
PWID	15	3.8
Unknown		
Place of residence		
Urban area	315	81.8
Rural	70	18.8
Unknown		
CDC disease stage		
A	175	45.5
B	93	24.1
C	117	30.4
Unknown		
OSTI		
Positive	87	22.6
Negative	298	77.4
Unknown		
Median age		
Male	34.2	
Female	39.5	

MSM, men who have sex with men; PWID, people who inject drugs; CDC, Center for Disease Control; OSTI, other sexually transmitted infections.

(**Figure 6**) and the heterosexual clades (**Figure 7**). Of note, for all three investigated MSM clades Re over 1 was present during the whole analyzed period (**Figures 6A–C**). Specifically, the highest Re (maximum value of median Re = 3.2) was found for the most expanded MSM transmission cluster, composed of 15 viral sequences, that also dated to the most recent period (**Figure 6A**). Similarly, for the second transmission cluster, Re was found to be over 1 during the whole period, reaching the maximum value of 2.3 in 2009 that remained constant for almost 5 years, followed by a decreasing phase, but still retaining value above 1 (**Figure 6B**). For the transmission network, Re was found to be slightly above 2 for almost 15 years, followed by an increasing phase in 2003, when it reached the value of 2.6 that remained constant approaching the present time (**Figure 6C**). On the other hand, birth-death skyline plot analyses for heterosexual subtype B monophyletic clade showed Re value below 1, suggesting inactivity for almost 10 years, followed by an active phase when it reached the maximum value of Re in the beginning of the decade of the 2000s, and then a decreasing phase with value below 1 thereafter (**Figure 7A**). A similar BDM skyline pattern was found for subtype C heterosexual clade, with the obtained Re (median estimates) over 1 for the time period 2003–2012 (**Figure 7B**).

Characterizing the Latent Classes

Akaike information criteria and the BIC values were used as indicated to select the number of latent classes in our LCA model. The best-fitting model calculated in R software in this research suggested five latent classes, and was identified by considering the lowest log likelihood of AIC and BIC values before these values increased with the addition of another class. **Figure 8** is a graphic presentation of AIC and BIC numerical data used as statistical support for the number of latent classes. The grouping of patients into five latent classes based on the distribution of different characteristics is shown in **Figure 9**. Transmission clusters were dominantly present in 2/5 classes. The first class included young MSM subjects aged up to 25 years old, living in Belgrade, with HIV infection diagnosed within the last 3 years of the study period, with 73% of coinfection with HBV and/or other STIs and the highest percentage (53%) of grouping within transmission clusters. The second class encompassed MSM subjects aged up to 45 years of age with high education, HIV diagnosis in the period from the mid-2000s onward, with less than 10% of coinfection with HBV and other STIs and around 20% of sequences found grouped in transmission clusters. In the further three classes no tendency to group in clusters was shown: the third latent class included equal percentages of the MSM transmission group and the heterosexual male subjects, diagnosed during the whole period of the HIV epidemic, from Belgrade, with high percentage of HBV but no other STIs; the fourth latent class encompassed heterosexual subjects of both sexes, aged over 25, with the date of the HIV diagnosis predominantly in the mid-2000s, mostly with secondary education, living in rural areas; the fifth class included female subjects reporting intravenous drug use (IVDU) as a risk for HIV infection, living in Belgrade, with very high percentage of HCV coinfection and diagnosed during whole period of HIV epidemic in Serbia.

DISCUSSION

Here, we present the study integrating modeling, phylogenetic and socio-demographic analyses of the HIV epidemic in Serbia, to elucidate features, dynamics, and trends of the local sub-epidemics.

The enormous genetic diversity of HIV-1 allows accurate patterns of evolution and transmission to be obtained from samples collected over a relatively short time (Pybus and Rambaut, 2009; Baele et al., 2017). Analysis of phylodynamics makes for a very useful tool for studying molecular epidemiology and aiding public health planning (Pybus and Rambaut, 2009; Baele et al., 2017; Vrancken et al., 2017). If these kinds of sophisticated analyses are based on a multidisciplinary approach, they may give a quantitative description of transmission networks, by identifying socio-demographic correlates of clustering: how virus lineages are restricted to or mix among, different demographic and epidemiological subgroups (Kostaki et al., 2017; Paraskevis et al., 2017).

In the current study, logistic growth modeling of the epidemic in Serbia was performed, based on the registered data on HIV incidence in the period 1984–2016. The obtained model was further used to predict the epidemic trends until 2030. The results for the general and MSM populations in Serbia imply early, exponential phase of HIV epidemic in both

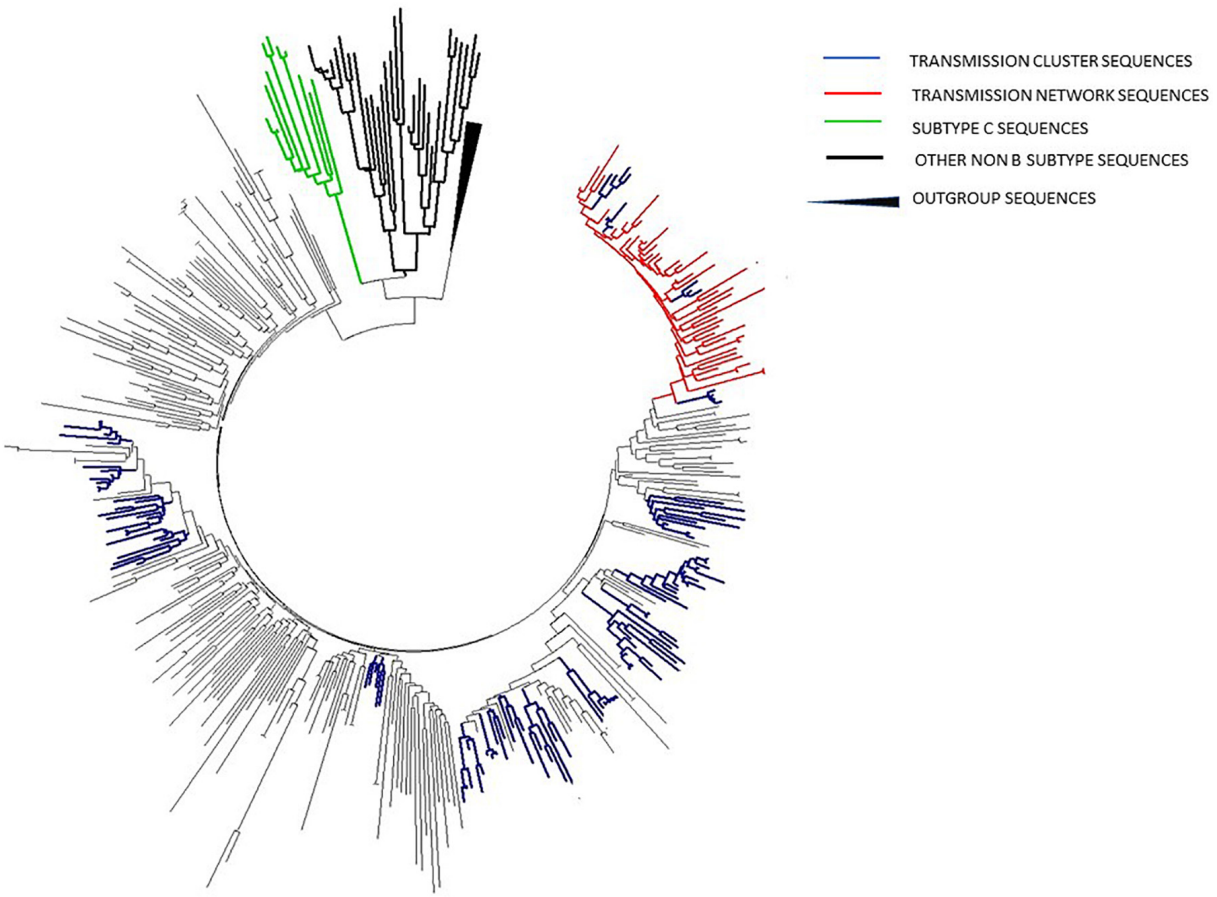


FIGURE 3 | ML phylogenetic tree constructed in MEGA software V6.1 including all sequences analyzed in the current study together with reference sequences and background sequences from the NCBI database.

cases, with a much steeper slope in the MSM population, potentially providing overall active epidemiological background for transmission clusters to occur. Conversely, in PWID and heterosexual transmission groups, plateaued growth was observed (**Supplementary Figures 4**, **5**). Accurate estimates are important to understand the true burden of HIV, the corresponding need for treatment, and for optimizing testing strategies for HIV. Incidence is the best measure for monitoring infections; however, it is difficult to obtain accurate estimates over time. One of the main approaches to reconstruct the HIV incidence curve is based on the reported number of HIV diagnoses. With the prevalence of HIV infection below 0.1%, Serbia is ranked as a low-prevalence country; however, it is of great importance to study the epidemiological situation in populations at risk such as the MSM and PWID transmission groups and among sex workers. Overall, the highest prevalence of HIV in Serbia has been estimated for MSM (8.3%), much higher when compared to the general population, as well as to other groups at risk such as sex workers (1.8%) and PWID transmission group (1.6%) (UNAIDS, 2017).

A further application of the model for a time period beyond 2017 estimated approaching the plateau in new HIV infections number by 2030 in general population, whereas in MSM transmission group an exponential-like curve with no signs of a plateau in the studied time period was seen. Potential disadvantages of the applied approach include a number of potential biases and difficulties, such as lack of data, underreporting, lack of information on changes over time, etc. However, the obtained results of logistic growth modeling are in line with recent data showing an increase in HIV incidence in Serbia. The incidence in 2000 was 0.02 per 1,000 adults, and in 2016 was 0.05 per 1,000 adults (UNAIDS, 2017). Estimated new HIV infections are decreasing globally, but increasing in the WHO European region, mostly in Eastern and Central Europe (ECDC, 2017; Mravčík et al., 2017). In Serbia, these findings might reflect the fact that though efforts have been made to decentralize preventive and advisory services, the coverage of high-risk groups with preventive services could still be insufficient. This may be due to the lack of recognition of risk behaviors or fear of further stigmatization, especially in smaller communities (Cousins, 2018). In Serbia, a much higher usage of condoms has been reported in professional sex workers than in MSM (UNAIDS, 2017). Moreover, within the MSM population, several vulnerable categories can be found: young men (especially

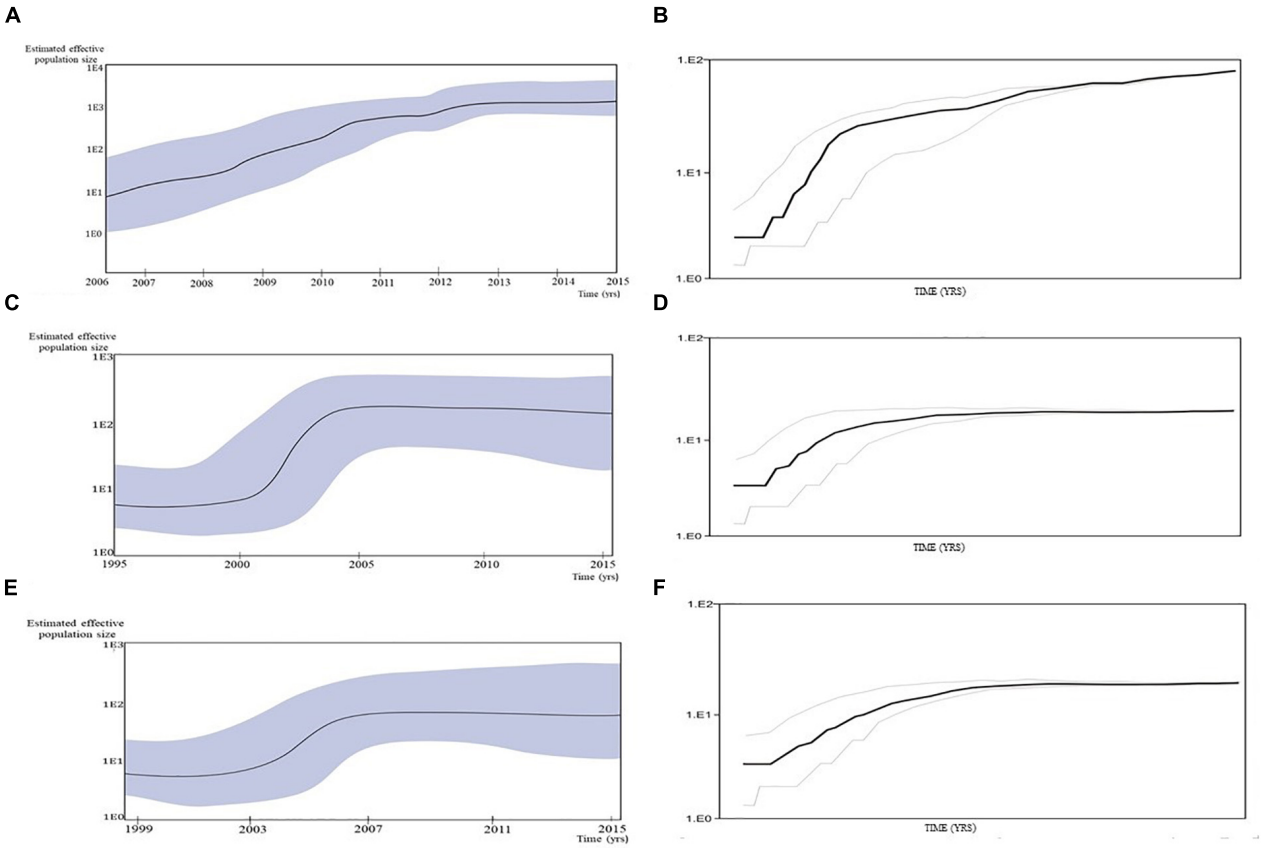


FIGURE 4 | Population growth and the cumulative number of lineages (infections) in a logarithmic scale over time for the **(A,B)** transmission cluster of 15 sequences; **(C,D)** transmission network; **(E,F)** transmission cluster of 11 sequences. The median estimate of the effective number of infections (solid line) and 95% confidence limits of the estimate (dashed lines) are shown in each graphic.

underage boys), MSM involved in sex work (which is illegal), and bisexual men (in particular men who define themselves as heterosexual but who have sex with other men). This latter group may be less likely to be informed than most gay men (Godinho et al., 2005).

The better understanding of genetic history, pattern, and dynamics of HIV transmission on a population level could lead to developments in prevention services, targeting factors associated with onward HIV-1 transmission (Granich et al., 2009; Cohen et al., 2011; Vrancken et al., 2017). In order to determine trends in the HIV epidemic in Serbia, based on HIV-1 pol sequence data generated through the antiretroviral resistance testing, we investigated the transmission clusters and phylodynamic profiles, using in-depth phylogenetic analyses that involve the ML and Bayesian coalescence strategy. These kinds of sophisticated analyses provide an insight into the epidemic trends and patterns of the evolutionary history of a certain viral population, revealing the size of transmission clusters and the dynamics of transmission within them. Interpretation of these results, however, may be hampered if the analyses are not associating other traits with clustering, which serve as determinants of the transmission dynamics and cluster activity (Brenner et al., 2013; Grabowski and Redd, 2014; Chan et al., 2015). Our previous study established the role of transmission clusters/networks in HIV epidemic spread in Serbia, however, phylodynamic aspects were not considered nor the impact of different socio-demographic and clinical characteristics on the clustering patterns (Siljic et al., 2013). In the current study, we found a high percentage of 53% of sequences involved in transmission clusters/network in Serbia. This finding is in line and even higher than our previous findings and studies of other epidemics (Leigh Brown et al., 2011; Siljic et al., 2013; Ragonnet-Cronin et al., 2016; Chaillon et al., 2017; Parczewski et al., 2017). Notably, it is considered that assessment of HIV transmission clusters is influenced by sampling density – the higher the sampling density, the more important the clustering (Novitsky et al., 2014). However, methods and definitions of transmission clusters in HIV phylogenies are still hotly debated – recent findings suggest that bootstrap support should be replaced or complemented by newer methods, such as refined, gradual function in large data sets (Lemoine et al., 2018). In view of the analyzed sample size we maintained a more conservative approach by using both bootstrap and Bayesian posterior probability, as the latter provides closer estimates of the true probabilities of recovering clades (Wilcox et al., 2002). Again, having in mind the studied sample size, the

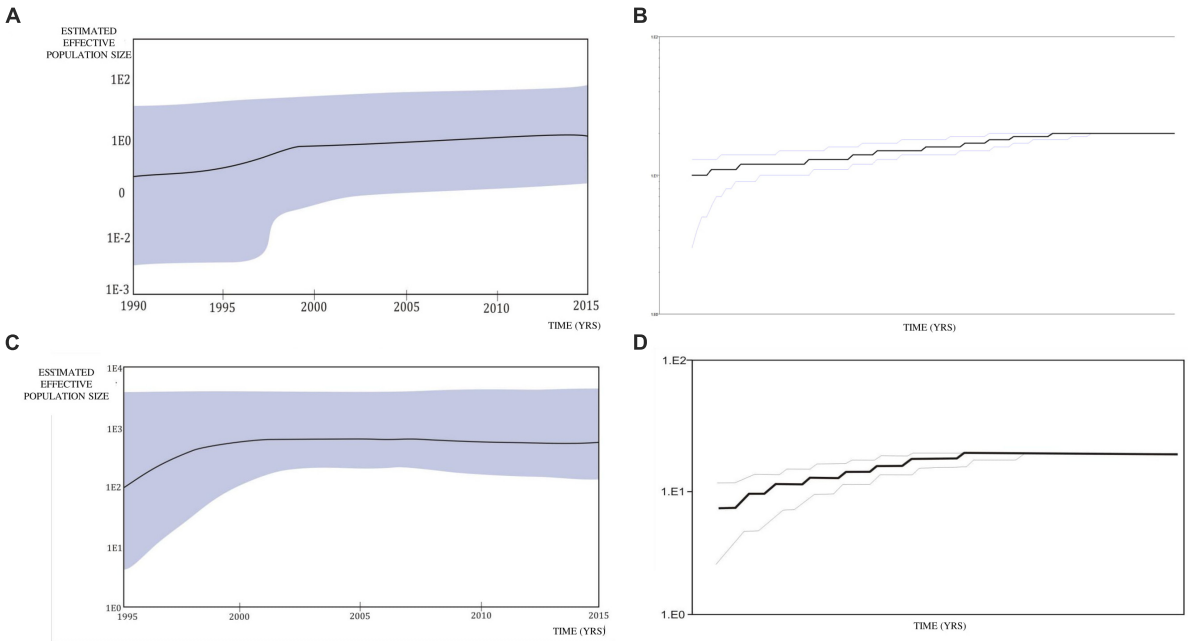


FIGURE 5 | Bayesian Skyline plot and logistic growth analyses performed in BEAST software package v1.7.5 presenting population growth and the cumulative number of lineages (infections) in a logarithmic scale over time for **(A)** and **(B)** the subtype B clade composed of sequences from heterosexuals; **(C)** and **(D)** the subtype C clade composed of sequences from heterosexuals.

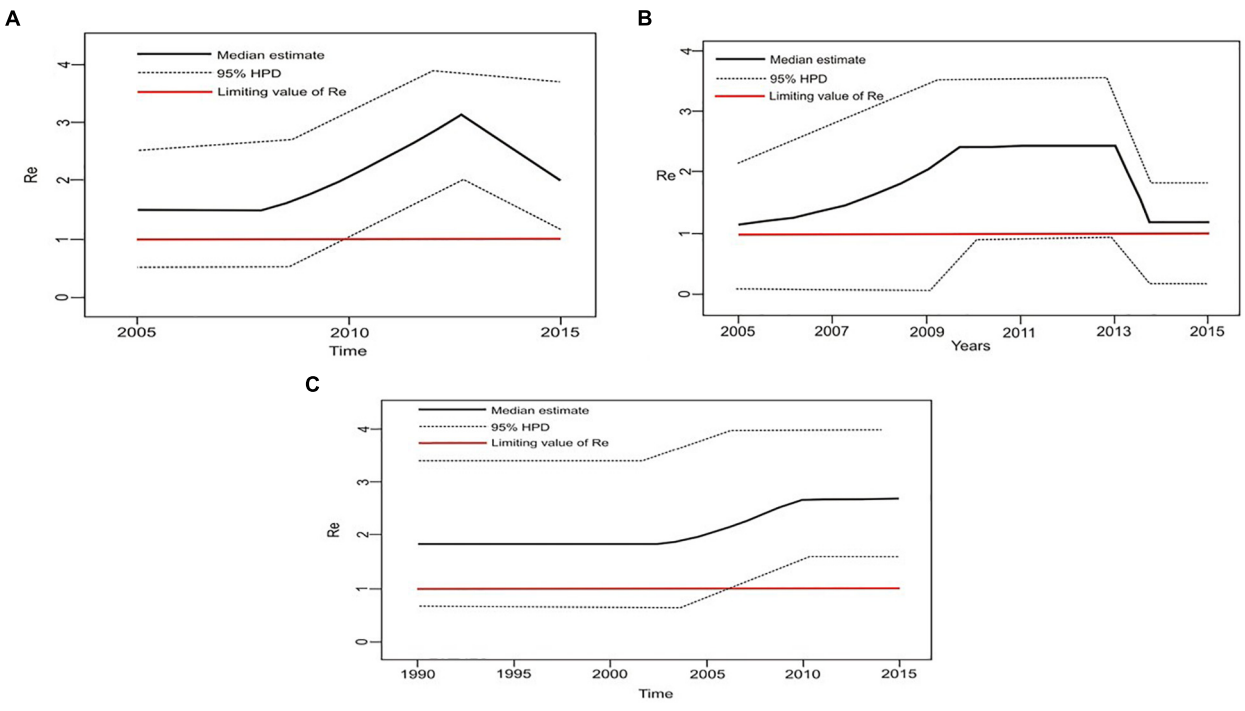


FIGURE 6 | The estimated birth-death skyline serial models by BEAST v2.1 presenting the effective reproductive number (Re) over time for the **(A)** transmission cluster of 15 sequences **(B)** transmission network **(C)** transmission cluster of 11 sequences.

obtained, high enough total percentage of 53% of clustered sequences, may be even considered as a rather conservative, lower edge estimate. Our results suggest that HIV spread is driven by local MSM transmission, as observed also in other

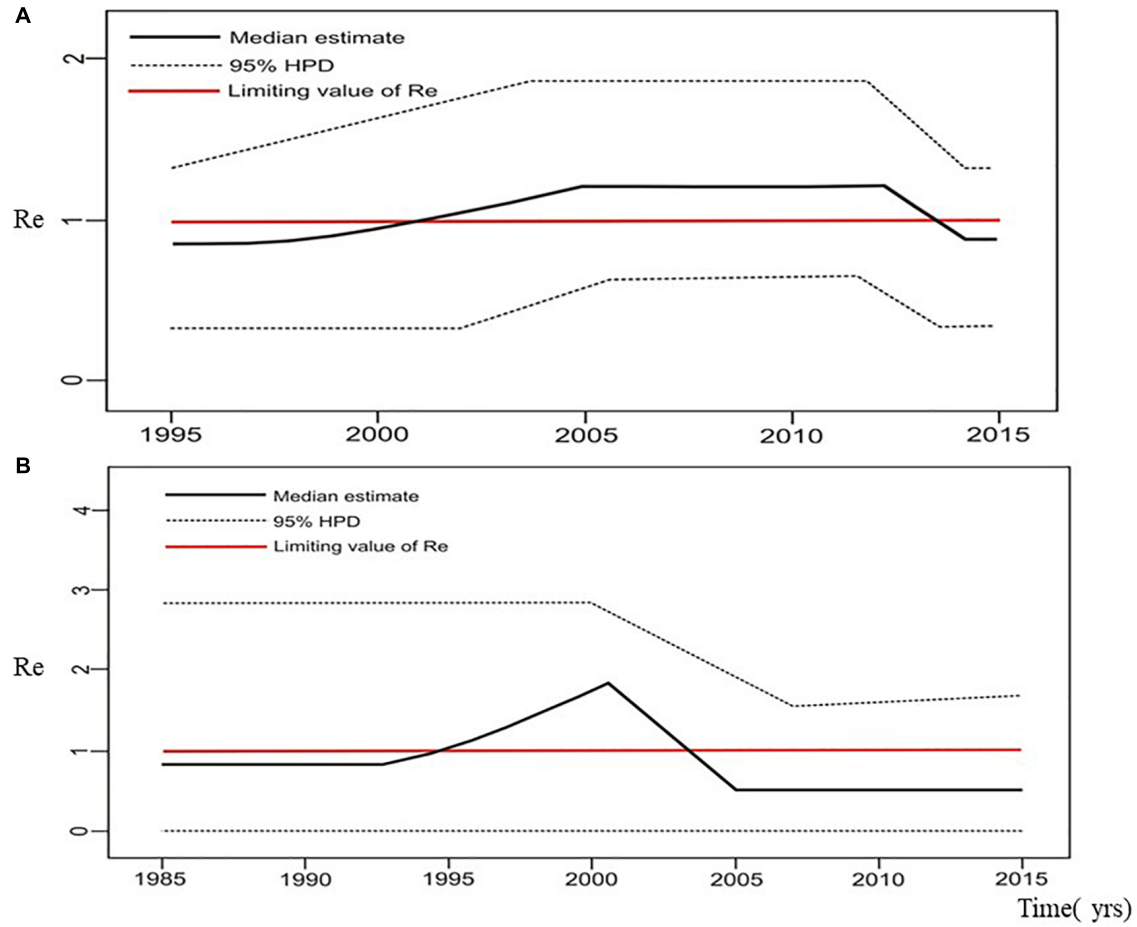


FIGURE 7 | The estimated birth-death skyline serial models by BEAST v2.1. presenting the effective reproductive number (Re) over time for the **(A)** subtype B clade composed of sequences from heterosexuals **(B)** subtype C clade composed of sequences from heterosexuals.

European countries (Sullivan et al., 2009; van Griensven et al., 2009; Frentz et al., 2013; Phillips et al., 2013). Similar to some other European countries, clustering sequences from MSM were found within subtype B (Abecasis et al., 2013; Esbjörnsson et al., 2016). Furthermore, we found an extension of previously reported small transmission clusters related to young newly diagnosed MSM patients (Siljic et al., 2013). Most of the heterosexual transmissions seemed to be limited to transmission pairs and small clusters, without substantial further spread of the infection. All things considered, our results show that local HIV transmission in Serbia is mainly driven by MSM transmission clusters.

Based on the effective reproductive number (Re) estimated through birth-death plots, together with the number of infections over time, significant differences between the MSM and heterosexual clades were found. MSM subtype B clades showed mean reproductive number over one during the whole investigated period, with the steepest slopes and a recent tMRCA, in line with a growing trend in the number of transmissions in years approaching the end of the study period. On the contrary, heterosexual clades in both studied subtypes – B and C, showed modest growth and stagnation. This finding could be influenced by the analyzed sample size, with a possible underestimation and bias toward being slower/lower at the moment compared to the MSM route but with possibility of change. However, neither epidemiological data and trends nor other analytical approaches (such as LCA analysis) imply that changes in the sub epidemic dynamics would be plausible.

The obtained results imply that none of the clusters will stop to be active in the near future as most clusters contain recent infections and are rejuvenated by the inclusion of younger men, and many have a reproduction number greater than the epidemic threshold. Although the transmission network was defined to be active, with reproductive potential striking higher than epidemic threshold and increasing in later years, it also appears to be "aging," while clusters appear to be of lower mean age; especially the most expanded cluster that was found to be "the youngest" was also found to be the one with the highest reproductive potential. On the other hand, particularly worrying is the fact that even though the transmission network comprised sequences from MSM diagnosed early in infection, this did not seem to have stopped clusters from growing. Other researchers showed

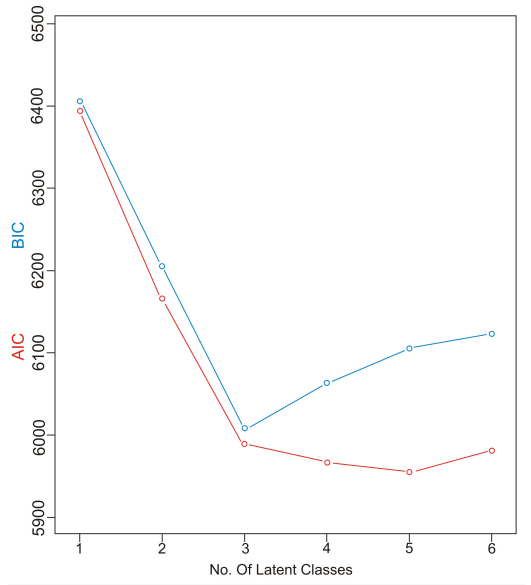


FIGURE 8 | AIC and BIC information criterion values used in model specification.

that a large part of onward transmission amongst MSM is related to the early phase of infection (Bezemer et al., 2010, 2015; Powers et al., 2011).

HIV transmission dynamics is considered to be shaped by a number of diverse constraints within the host, upon a transmission process and at the population level, influencing viral evolution (within-host and inter-host), virus and host genetics, complex interplay of between-host interactions but also important social and demographic factors (Theys et al., 2018). Hence, the inference of HIV-1 transmission dynamics and factors influencing epidemic spread may provide an important input for the design of efficient public health interventions, and new approaches are constantly being explored (Poon, 2016; Le Vu et al., 2018). The LCA analysis provided further insights by combining clustering patterns and a set of characteristics associated with patients such as socio-demographic, clinical, transmission risk, diagnosis date. Together, these features allow assessing multiple determinants of the local HIV transmission, helping to understand its occurrence in a "real life" manner. This analysis has separated five subgroups (latent classes) with various combinations of analyzed characteristics that affect differently the epidemic in Serbia. Importantly, four factors were found to be significantly associated with patients belonging to clusters: very young age, residing in the capital city, recent diagnosis and a high rate of STI and HBV coinfections.

According to the latest European epidemiological data, sex between men remains the predominant mode of HIV transmission and MSM are disproportionately at risk for and affected by HIV, STIs and viral hepatitis (ECDC, 2015, 2017, 2018). A similar situation has been increasingly described in different regions worldwide (Qi et al., 2015). Current WHO guidelines recommend oral pre-exposure prophylaxis (PrEP) to be offered as an additional prevention choice for people at substantial risk of HIV infection as part of a combination of HIV prevention approaches, where 'substantial risk' of HIV infection is provisionally defined as HIV incidence greater than 3 per 100 person-years (WHO, 2015). In Serbia, WHO 90–90–90 target has been embraced, however, PrEP is not yet available. Monitoring the HIV epidemic is essential for assessing the impact of effective HIV prevention interventions, determining public health priorities, and estimating current and future health care needs.

In summary, the results presented imply that the HIV epidemic in Serbia is still in the exponential growth phase, in particular, related to the MSM transmission that is estimated to retain the steep growth curve until 2030. The previously described tendency of cluster formation in this group has been confirmed. The obtained results imply that an average new HIV patient in Serbia is a young man with concomitant STIs. Together, these findings provide a useful insight that may prove to be vital for prospective public health priorities and interventions, in particular, relative to the 90–90–90 targets and considerations of PrEP.

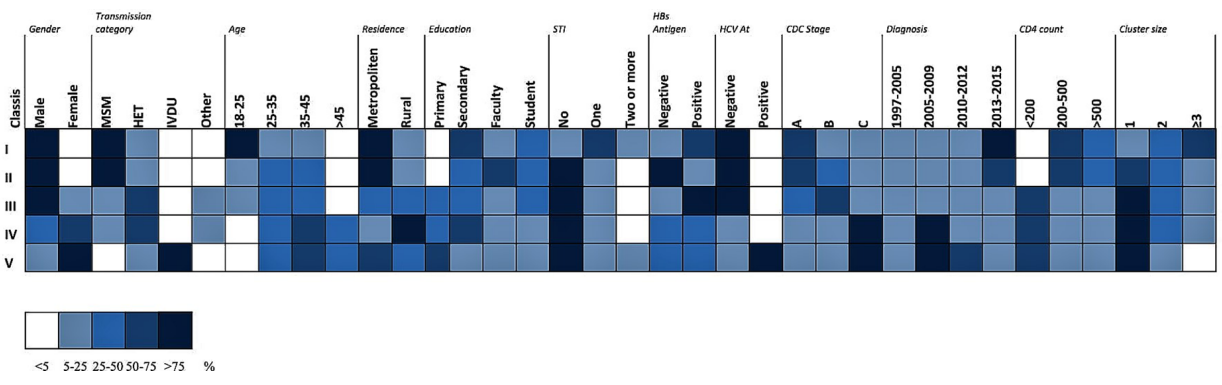


FIGURE 9 | LCA performed in R software with a poLCA software package on data of HIV-1 infected individuals included in the current research. The distribution of participant characteristics within the latent class and predicted proportions in each cluster size by latent class are shown. Darker colors represent higher proportion as shown in the legend.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of University of Belgrade Faculty of Medicine Ethical Committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Belgrade Faculty of Medicine Ethical Committee, decision No 29/V-11.

AUTHOR CONTRIBUTIONS

MS, LJ, and MŠ conceived and designed the study. MŠ, VĆ, and MT performed molecular analyses. DS, IP-P, JR, and DJ collected and analyzed the epidemiological data. LJ performed modeling analysis. LJ, MŠ, and VĆ performed phylogenetic analyses. MS, LJ, and MŠ have been involved in drafting the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was partly supported by the grant OI175024, from the Ministry of Education, Science and Technological Development, Serbia.

ACKNOWLEDGMENTS

We thank Sretko Luković for his contribution to the initial idea of the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00287/full#supplementary-material

REFERENCES

Abecasis, A. B., Wensing, A. M., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A., et al. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 10:7. doi: 10.1186/1742-4690-10-7

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the 2nd International Symposium on Information Theory*, eds B. N. Petrov and F. Csaki (Budapest: Akademiai Kiado), 267–281.

Baele, G., Suchard, M. A., Rambaut, A., and Lemey, P. (2017). Emerging concepts of data integration in pathogen phylodynamics. *Syst. Biol.* 66, e47–e65. doi: 10.1093/sysbio/syw054

Bezemer, D., Cori, A., Ratmann, O., van Sighem, A., Hermanides, H. S., Dutilh, B. E., et al. (2015). Dispersion of the HIV-1 epidemic in men who have sex with men in the netherlands: a combined mathematical model and phylogenetic analysis. *PLoS Med.* 12:e1001898. doi: 10.1371/journal.pmed.1001898

Bezemer, D., van Sighem, A., Lukashov, V. V., van der Hoek, L., Back, N., Schuurman, R., et al. (2010). Transmission networks of HIV-1 among men having sex with men in the Netherlands. *AIDS* 24, 271–282. doi: 10.1097/QAD.0b013e328333ddee

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Brenner, B., Wainberg, M. A., and Roger, M. (2013). Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. *AIDS* 27, 1045–1057. doi: 10.1097/QAD.0b013e32835cffd9

Chaillon, A., Essat, A., Frange, P., Smith, D. M., Delaugerre, C., Barin, F., et al. (2017). Spatiotemporal dynamics of HIV-1 transmission in France (1999–2014) and impact of targeted prevention strategies. *Retrovirology* 14:15. doi: 10.1186/s12977-017-0339-4

Chan, P. A., Hogan, J. W., Huang, A., DeLong, A., Salemi, M., Mayer, K. H., et al. (2015). Phylogenetic investigation of a statewide HIV-1 epidemic reveals ongoing and active transmission networks among men who have sex with men. *J. Acquir. Immune Defic. Syndr.* 70, 428–435. doi: 10.1097/QAI.0000000000000786

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., et al. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med.* 365, 493–495. doi: 10.1056/NEJMoa1105243

Cousins, S. (2018). HIV in Serbia: stigma and a stagnant HIV response. *Lancet HIV* 5, e343–e344. doi: 10.1016/S2352-3018(18)30144-9

Drummond, A., Pybus, O. G., and Rambaut, A. (2003). Inference of viral evolutionary rates from molecular sequences. *Adv. Parasitol.* 54, 331–358.

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

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

ECDC (2015). *European Centre for Disease Prevention and Control. Annual Epidemiological Report 2014 - Sexually Transmitted Infections, Including HIV and Blood-borne Viruses*. Available at: https://ecdc.europa.eu/en/publications-data/sexually-transmitted-infections-including-hiv-and-blood-borne-viruses-annual.

ECDC (2017). *European Centre for Disease Prevention and Control/WHO Regional Office for Europe. HIV/AIDS Surveillance in Europe 2017 – 2016 data*. Available at: https://ecdc.europa.eu/en/publications-data/hivaids-surveillance-europe-2017-2016-data.

ECDC (2018). *European Centre for Disease Prevention and Control. Hepatitis B and C Epidemiology in Selected Population Groups in the EU/EEA*. Stockholm: ECDC.

Esbjörnsson, J., Mild, M., Audelin, A., Fonager, J., Skar, H., Bruun Jørgensen, L., et al. (2016). HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. *Virus Evol.* 2:vew010. doi: 10.1093/ve/vew010

Fonseca, M. G. P., and Bastos, F. I. (2007). Twenty-five years of the AIDS epidemic in Brazil: principal epidemiological findings, 1980–2005. *Rev. Saude Publica* 23, S333–S343. doi: 10.1590/S0102-311X2007001500002

Frentz, D., Wensing, A. M., Albert, J., Paraskevis, D., Abecasis, A. B., Hamouda, O., et al. (2013). Limited cross-border infections in patients newly diagnosed with HIV in Europe. *Retrovirology* 10:36. doi: 10.1186/1742-4690-10-36

Godinho, J., Jaganjac, N., Eckertz, D., Renton, A., Novotny, T., and Garbus, L. (2005). *HIV/AIDS in the Western Balkans Priorities for Early Prevention in a High-Risk Environment*. Washington, D.C: World Bank. doi: 10.1596/0-8213-6394-8

Grabowski, M. K., and Redd, A. D. (2014). Molecular tools for studying HIV transmission in sexual networks. *Curr. Opin. HIV AIDS* 9, 126–133. doi: 10.1097/COH.0000000000000040

Granich, R. M., Gilks, C. F., Dye, C., De Cock, K. M., and Williams, B. (2009). Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet* 373, 48–57. doi: 10.1016/S0140-6736(08)61697-9

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Kostaki, E., Magiorkinis, G., Psichogiou, M., Flampouris, A., Iliopoulos, P., Papachristou, E., et al. (2017). Detailed molecular surveillance of the HIV-1 outbreak among people who inject drugs (PWID) in athens during a period of four years. *Curr. HIV Res.* 15, 396–404. doi: 10.2174/1570162X15666171120104048

Le Vu, S., Ratmann, O., Delpech, V., Brown, A. E., Gillc, O. N., Tostevin, A., et al. (2018). Comparison of cluster-based and source-attribution methods for estimating transmission risk using large HIV sequence databases. *Epidemics* 23, 1–10. doi: 10.1016/j.epidem.2017.10.001

Leigh Brown, A. J., Lycett, S. J., Weinert, L., Hughes, G. J., Fearnhill, E., Dunn, D. T., et al. (2011). Transmission network parameters estimated from HIV sequences fora nationwide epidemic. *J. Infect. Dis.* 204, 1463–1469. doi: 10.1093/infdis/jir550

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing felsenstein phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-0.)

Linzer, D. A., and Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *J. Stat. Softw.* 42. doi: 10.18637/jss.v042.i10

Magiorkinis, G., Angelis, K., Mamais, I., Katzourakis, A., Hatzakis, A., Albert, J., et al. (2016). The global spread of HIV-1 subtype B epidemic. *Infect. Genet. Evol.* 46, 169–179. doi: 10.1016/j.meegid.2016.05.041

Mravćík, V., Pitoòák, M., Hejzák, R., Janíková, B., and Procházka, I. (2017). HIV epidemic among men who have sex with men in the Czech Republic 2016: high time for targeted action. *Eur. Surveill.* 22, 17–00079. doi: 10.2807/1560-7917.ES.2017.22.48.17-00079

Novitsky, V., Moyo, S., Lei, Q., DeGruttola, V., and Essex, M. (2014). Impact of sampling density on the extent of HIV clustering. *AIDS Res. Hum. Retrovir.* 30, 1226–1235. doi: 10.1089/aid.2014.0173

Paraskevis, D., Kostaki, E., Gargalianos, P., Xylomenos, G., Lazanas, M., Skoutelis, A., et al. (2017). Transmission dynamics of HIV-1 drug resistance among treatment-naïve individuals in greece: the added value of molecular epidemiology to public health. *Genes* 8:E322. doi: 10.3390/genes8110322

Parczewski, M., Leszczyszyn-Pynka, M., Witak-Jędra, M., Szetela, B., Gąsiorowski, J., Knysz, B., et al. (2017). Expanding HIV-1 subtype B transmission networks among men who have sex with men in Poland. *PLoS One* 12:e0172473. doi: 10.1371/journal.pone.0172473

Phillips, A. N., Cambiano, V., Nakagawa, F., Brown, A. E., Lampe, F., Rodger, A. et al. (2013). Increased HIV incidence in men who have sex with men despite high levels of ART-induced viral suppression: analysis of an extensively documented epidemic. *PLoS One* 8:e55312. doi: 10.1371/journal.pone.0055312

Poon, A. F. Y. (2016). Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus Evol.* 2:vew031. doi: 10.1093/ve/vew031

Posada, D. (2009). Selection of models of DNA evolution with jModelTest. *Methods Mol. Biol.* 537, 93–112. doi: 10.1007/978-1-59745-251-9_5

Powers, K. A., Ghani, A. C., Miller, W. C., Hoffman, I. F., Pettifor, A. E., Kamanga, G., et al. (2011). The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. *Lancet* 378, 256–268. doi: 10.1016/S0140-6736(11)60842-8

Pybus, O. G., and Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10, 540–550. doi: 10.1038/nrg2583

Qi, J., Zhang, D., Fu, X., Li, C., Meng, S., Dai, M., et al. (2015). High risks of HIV transmission for men who have sex with men — a comparison of risk factors of HIV infection among MSM associated with recruitment channels in 15 Cities of China. *PLoS One* 10:e0121267. doi: 10.1371/journal.pone.0121267

Ragonnet-Cronin, M., Lycett, S. J., Hodcroft, E. B., Hue, S., Fearnhill, E., Brown, A. E., et al. (2016). Transmission of Non-B HIV subtypes in the United Kingdom is increasingly driven by large non-heterosexual transmission clusters. *J. Infect. Dis.* 213, 1410–1418. doi: 10.1093/infdis/jiv758

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–56. doi: 10.1126/science.288.5463.55d

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Schreiber, J. B. (2017) Latent class analysis: an example for reporting results. *Res. Social. Adm. Pharm.* 13, 1196–1201. doi: 10.1016/j.sapharm.2016.11.011

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Sherrod, P. H. (2010). *NLREG Nonlinear Regression Analysis Program*. Available at: http://www.nlreg.com/index.htm

Siljic, M., Salemovic, D., Cirkovic, V., Pesic-Pavlovic, I., Ranin, J., Todorovic, M., et al. (2017). Forensic application of phylogenetic analyses - exploration of suspected HIV-1 transmission case. *Forensic Sci. Int. Genet.* 27, 100–105. doi: 10.1016/j.fsigen.2016.12.006

Siljic, M., Salemovic, D., Jevtovic, D., Pesic-Pavlovic, I., Zerjav, S., Nikolic, V., et al. (2013). Molecular typing of the local HIV-1 epidemic in Serbia. *Infect. Genet. Evol.* 19, 378–385. doi: 10.1016/j.meegid.06.015

Snoeck, J., Riva, C., Steegen, K., Schrooten, Y., Maes, B., Vergne, L., et al. (2005). Optimization of a genotypic assay applicable to all human immunodeficiency virus type 1 protease and reverse transcriptase subtypes. *J. Virol. Methods* 128, 47–53. doi: 10.1016/j.jviromet.2005.04.001

Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., et al. (2012). Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* 29, 347–357. doi: 10.1093/molbev/msr217

Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. U.S.A.* 110, 228–233. doi: 10.1073/pnas.1207965110

Stanojevic, M., Papa, A., Papadimitriou, E., Zerjav, S., Jevtovic, D., Salemovic, D. et al. (2002) HIV-1 subtypes in Yugoslavia. *AIDS Res. Hum. Retroviruses* 18, 519–522. doi: 10.1089/088922202317406673

Stanojevic, M., Siljic, M., Salemovic, D., Pesic-Pavlovic, I., Zerjav, S., Nikolic, V., et al. (2014). Ten years survey of primary HIV-1 resistance in Serbia: the occurrence of multiclass resistance. *AIDS Res. Hum. Retrovir.* 30, 634–641. doi: 10.1089/AID.2013.0270

Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* 18, 1001–1013. doi: 10.1093/oxfordjournals.molbev.a003872

Sullivan, P. S., Hamouda, O., Delpech, V., Geduld, J. E., Prejean, J., Semaille, C., et al. (2009). Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996–2005. *Ann. Epidemiol.* 19, 423–431. doi: 10.1016/j.annepidem.2009.03.004

Swofford, D. L. (1999). *PAUP* 40: Phylogenetic Analysis Using Parsimony (*and other methods), version 40b10*. Sunderland, MA: Sinauer Associates Inc.

Theys, K., Libin, P., Pineda- Pena, A.-C., Nowe, A., Vandamme, A.-M., and Abecasis, A. B. (2018). The impact of HIV-1 within host evolution on transmission dynamics. *Curr. Opin. Virol.* 28, 92–101. doi: 10.1016/j.coviro.2017.12.001

UNAIDS (2017). *Ending AIDS: Progress Towards the 90-90-90 targets. Global AIDS update 2017*. Available at: http://www.unaids.org/en/resources/documents/2017/20170720_Global_AIDS_update_2017

van Griensven, F., de Lind van Wijngaarden, J. W., Baral, S., and Grulich, A. (2009). The global epidemic of HIV infection among men who have sex with men. *Curr. Opin. HIV AIDS* 4, 300–307. doi: 10.1186/1742-4690-10-7

Vandermeer, J. (2010). How populations grow: the exponential and logistic equations. *Nat. Educ. Knowl.* 3:15.

Vrancken, B., Adachi, D., Benedet, M., Singh, A., Read, R., Shafran, S., et al. (2017). The multi-faceted dynamics of HIV-1 transmission in Northern Alberta: a

combined analysis of virus genetic and public health data. *Infect. Genet. Evol.* 52, 100–105. doi: 10.1016/j.meegid.2017.04.005

Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., et al. (2017). UpdateoftheDrug ResistanceMutations in HIV-1. *Top. Antivir. Med.* 4, 132–133.

WHO (2015). *Guideline on When to Start Antiretroviral Therapy and on Pre-exposure Prophylaxis for HIV*. Geneva: World Health Organization.

Wilcox, T. P., Zwickl, D. J., Heath, T. A., and Hillis, D. M. (2002). Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25, 361–371. doi: 10.1016/S1055-7903(02)00244-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 February 2019
doi: 10.3389/fmicb.2019.00320

Human Immunodeficiency Virus-1 Diversity in the Moscow Region, Russia: Phylodynamics of the Most Common Subtypes

Aleksey Lebedev[1], *Natalya Lebedeva*[2], *Fedor Moskaleychik*[1], *Alexander Pronin*[2], *Elena Kazennova*[1] *and Marina Bobkova*[1]*

[1] Laboratory of T-Lymphotropic Viruses, N.F. Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia, [2] Moscow Regional AIDS Centre, Moscow, Russia

This study analyzes the HIV-1 subtype diversity and its phylodynamics in Moscow region, which is the most densely populated area of Russia characterized by high rates of internal and external migration. The demographic and viral data from 896 HIV-infected individuals collected during 2011–2016 were analyzed. The study revealed broad diversity in the HIV-1 subtypes found in Moscow, which included A6 (85.1%), B (7.6%), CRF02_AG (1.2%) and URF_A6/B recombinants (4.2%). Other HIV-1 subtypes were detected as single cases. While A6 was most prevalent (>86.0%) among heterosexuals, injecting drug users and cases of mother-to-child transmission of HIV, subtype B (76.3%) was more common in men who have sex with men. Phylogenetic reconstruction revealed that the A6 sequences were introduced into the epidemic cluster that arose approximately around 1998. Within the subtype B, six major epidemic clusters were identified, each of which contained strains associated with only one or two dominant transmission routes. The date of origin of these clusters varied between 1980 and 1993, indicating that the HIV-1 B epidemic began much earlier than the HIV-1 A6 epidemic. Reconstruction of the demographic history of subtypes A6 and B identified at least two epidemic growth phases, which included an initial phase of exponential growth followed by a decline in the mid/late 2010s. Thus, our results indicate an increase in HIV-1 genetic diversity in Moscow region. They also help in understanding the HIV-1 temporal dynamics as well as the genetic relationships between its circulating strains.

Keywords: HIV-1 subtypes, diversity, phylodynamics, epidemic, TMRCA, transmission route, Moscow region, Russia

OPEN ACCESS

Edited by:
Michael M. Thomson,
Instituto de Salud Carlos III, Spain

Reviewed by:
Dimitrios Paraskevis,
National and Kapodistrian University of Athens, Greece
Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands

***Correspondence:**
Marina Bobkova
mrbobkova@mail.ru

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *12 October 2018*
Accepted: *06 February 2019*
Published: *26 February 2019*

Citation:
Lebedev A, Lebedeva N, Moskaleychik F, Pronin A, Kazennova E and Bobkova M (2019) Human Immunodeficiency Virus-1 Diversity in the Moscow Region, Russia: Phylodynamics of the Most Common Subtypes. Front. Microbiol. 10:320. doi: 10.3389/fmicb.2019.00320

INTRODUCTION

HIV-1 is a rapidly evolving human pathogen, characterized by significant genetic diversity. The pandemic group M is subdivided into nine subtypes (A–D, F–H, J, K), each of which have been further divided into sub-subtypes (A1-A4, A6, F1-F2), and multiple inter-subtype recombinants (CRFs) (Foley et al., 2016). The global distribution of HIV subtypes is heterogeneous and varies significantly even within the same continent/country (Hemelaar et al., 2011; Bobkova, 2013).

The HIV epidemic started in Russia in the mid-1990s with an outbreak of the subtype A infection among intravenous drug users (IDUs) (Bobkov et al., 1998). As per the current

nomenclature, the viral variant that caused the first epidemic and accounted for up to 90% of the infections is the sub-subtype A6 (formerly FSU-A or IDU-A) (Foley et al., 2016). During the following years, the HIV epidemic spread further affecting heterosexuals (HSX) and men who have sex with men (MSM). Ten years later sexual transmission accounted for about 37.3% of the new cases (FedAC, 2010).

With a shift in the dominant transmission route in Russia, the ratio of HIV subtypes also changed. The overall proportion of the non-A6-subtype increased from 7.0% in 2000 to 10.0–20.0% in 2010 (Bobkov et al., 1998; Thomson et al., 2009; Gashnikova et al., 2011). In the mid-2010s the prevalence of HIV-1 subtypes in Russia was as follows: A6 (∼70.0%), B (∼10.0%) and CRF063_02A1 (∼7.0%) (Lapovok et al., 2017). The subtype B was predominant (82%) among MSM (Dukhovlinova et al., 2015; Kazennova et al., 2017). The sub-subtype A6 was dominant in all territories of Russia, except the Siberian region and bordering Central Asia, where CRF063_02A1 and CRF02_AG recombinant variants were highly prevalent (Lapovok et al., 2014; Gashnikova et al., 2017; Kostaki et al., 2018).

During the last few years, the recombinant CRF063_02A1 has been responsible for more than 71.0% of the new HIV infections among HSXs and IDUs in Central Siberia (Gashnikova et al., 2015). Such changes in the prevalence of HIV variants in Russia are likely to affect the state of infection in Moscow region, which is characterized by a high population density and high rates of internal and external migration. Moscow located in the European part of Russia is a highly developed transport hub, for communications both within and outside Russia. The first HIV infection in Moscow region (not Moscow city) was reported in 1994; by the end of 2018 the number of people living with HIV (PLHIV) in this area was estimated to be 41,949 and the prevalence among adults has reached 0.7%.

There have been few studies on HIV diversity in Moscow region. While sub-subtype A6 has been described as the causative agent of the first HIV outbreak among IDUs probably introduced from St. Petersburg (Bobkov et al., 2001; Diez-Fuertes et al., 2015), by 2010 no significant changes were reported in the spectrum of circulating HIV subtypes [A6 (93.4%) and B (6.6%)] in the region (Giliazova et al., 2010). However, to date, there has been no information regarding the evolutionary history and population dynamics of HIV in Moscow region.

In this study, we aim to analyze the HIV-1 genetic diversity and reconstruct the temporal dynamics of its most common subtypes in Moscow region, using phylogenetic and phylodynamic approaches.

MATERIALS AND METHODS

Study Population

The study involved 896 patients who were diagnosed with HIV infection and received antiretroviral therapy (ART) between 2011 and 2016 at the Moscow RCAC which is 2.1% of PLHIV in this area. Three criteria for formation of study population have been applied: (1) well-documented history of HIV-infection in patients, (2) permanent residency in Moscow region, and (3) informed consent for the research study.

Genotyping and Dataset

Plasma samples were genotyped using the ViroSeq HIV-1 Genotyping System (Abbott, USA). Multiple sequence alignments were made using MAFFT (Katoh et al., 2017)The total length of the alignment was 1299 nucleotides which covered the entire protease and partial reverse transcriptase (positions 2253–3551, HXB2-numbering). HIV- 1 subtyping was performed using the COMET HIV-1 (Struck et al., 2014) and REGA Subtyping Tool v3.0 (Pineda-Pena et al., 2013) and subsequently confirmed by phylogenetic analysis. New inter-subtype or inter-CRF sequences were analyzed using jpHMM (Schultz et al., 2012). Moscow region sequences for A6 ($N = 5$) and B ($N = 9$) (2007–2008 sampling years) from the Los Alamos HIV-1 database (https://www.hiv.lanl.gov) were additionally included in the analysis. The epidemiological clusters were identified in the initial dataset of A6 ($N = 768$) and B ($N = 77$) subtypes. The evolutionary population dynamics was estimated in the optimized dataset by excluding (1) presence of more than 1% of ambiguous bases, (2) sequences with especially high evolutionary rates and (3) potential late chronic infections by AIDS symptoms. Finally, 320 and 71 sequences for A6 and B subtypes, respectively were selected. All codons associated with major drug-resistance mutations (Johnson et al., 2011) were removed from the final sequence alignment.

Phylogenetic Reconstruction and Phylodynamic Analysis

Selecting the best-fit model of nucleotide substitution was performed using jModelTest v.2.1.4 (Darriba et al., 2012). According to the Akaike information criterion (AIC), the best model for each dataset was General Time Reversible model with proportion of invariable sites and gamma-distributed rate variation among sites (GTR+I+G).The epidemiological clusters were identified using the maximum-likelihood phylogenetic analysis implemented by the IQ-TREE (Nguyen et al., 2015) with 1000 replicates for bootstrap and Shimodaira–Hasegawa (SH)-aLRT test. The clusters with an SH-aLRT support >0.9 were considered reliable.

The time of the most recent common ancestor (TMRCA) and the effective population size was estimated employing the Bayesian Markov Chain Monte Carlo (MCMC) approach using the BEAST v1.10.0 (Suchard et al., 2018). The temporal scale of the evolutionary process was inferred from the sampling dates of the sequences using a strict molecular clock model and the Bayesian Skyline coalescent tree prior. Comparing molecular clock models (as well as demographic models) in Bayesian framework were performed by calculating the AIC for MCMC with Tracer v1.7.1. Two independent MCMC chains were run of 50–500 $\times$ 10^6 steps with logging every 2000 generations, excluding first 25%. Convergence of the chains was estimated based on the Effective Sample Size in Tracer v1.7.1. The parameter estimates with ESS > 200 were accepted. We summarized the maximum clade credibility trees from the posterior distribution of trees using Tree Annotator and visualized them in FigTree v1.4.0. The datasets were analyzed with

TempEst (Rambaut et al., 2016) for evaluating the temporal signal sufficiency.

Statistical Analysis

Statistical analysis was performed using STATISTICA v10.0 (StatSoft, USA) with discrete categorical data and the Pearson's chi-squared (or Fisher's extact) test. Differences were considered significant at $P < 0.05$.

RESULTS

Study Population

We evaluated 896 HIV-1 infected patients with a median age of 35 years (range 1–67). The proportion of men and women was about the same (58.5 and 41.5%, respectively). The dominant transmission routes were heterosexual contact (49.8%), intravenous drug use (39.8%), mother-to-child transmission (MTCT, 5.5%) and MSM contact (4.2%). The demographic characteristics of patients grouped by subtypes are summarized in **Table 1**.

HIV Subtypes

Analysis of the pol sequences from HIV-1 infected patients confirmed the expected broad spectrum of viral subtypes in Moscow region (**Table 1**). These included 763 viruses (85.1%) classified as sub-subtype A6, 68 (7.6%) as subtype B, 11 (1.2%) as CRF02_AG and 16 (1.8%) as ≪other≫ subtypes. In addition to ≪pure≫ HIV-1 subtypes (or CRFs), 43 (4.8%) URFs containing genome segments of different subtypes were found. The sub-subtype A6 and CRF02_AG were found mainly in HSXs (50.3% and 72.7%, respectively) and IDUs (42.3 and 27.3%, respectively). Subtype B infected predominantly men (89.7%), through MSM (42.6%) and HSX (41.2%) exposures. Seven other HIV-1 subtypes were also identified.

Based on sequences identified by BLAST searches and phylogenetic analysis, sub-subtype A1 was found to be related to the East African strains (Uganda/Rwanda), subtype F1 to Angola and Romania, subtype C to the India/China C-strains; subtype G to the Portuguese/ Spanish G-strains and CRF01_AE to the Philippines AE-strains. Sub-subtype A6, CRF03_AB, CRF63_02A1, and CRF02_AG viruses were related to strains circulating in the former Soviet Union (FSU) countries. While most of the subtype B sequences (N=64) belonged to the Pandemic (≪Western-B≫) variant, four sequences belonged to the FSU-B variant, previously identified among IDUs in Ukraine (Nabatov et al., 2002) (**Figure 1**).

TABLE 1 | General characteristics of HIV-1 infected patients from Moscow region classified on subtypes.

		HIV-1 subtypes					*P*-value[b]
	Total (*N* = 896, 100.0%)	**A6 (*N* = 763, 85.1%)**	**B (*N* = 68, 7.6%)**	**URF_ A6/B (*N* = 38, 4.2%)**	**CRF02_AG (*N* = 11, 1.2%)**	**Others[a] (*N* = 16, 1.8%)**	
Age (years)							0.883[c]
< 30	170 (19.0)	143 (84.1)	15 (8.8)	7 (4.1)	1 (0.6)	4 (2.4)	
30–35	332 (37.0)	289 (87.1)	21 (6.3)	13 (3.9)	5 (1.5)	4 (1.2)	
> 35	394 (44.0)	331 (84.0)	32 (8.1)	18 (4.6)	5 (1.3)	8 (2.0)	
Gender							<0.001[c]
Male	524 (58.5)	424 (80.9)	61 (11.6)	23 (4.4)	4 (0.8)	12 (2.3)	
Female	372 (41.5)	339 (91.1)	7 (1.9)	15 (4.0)	7 (1.9)	4 (1.1)	
Risk group							0.001[d]
HSX	446 (49.8)	384 (86.1)	28 (6.3)	17 (3.8)	8 (1.8)	9 (2.0)	
MSM	38 (4.2)	6 (15.8)	29 (76.3)	1 (2.6)	–	2 (5.3)	
IDUs	357 (39.8)	323 (90.5)	11 (3.1)	17 (4.8)	3 (0.8)	3 (0.8)	
MTCT	49 (5.5)	44 (89.8)	–	3 (6.1)	–	2 (4.1)	
Unknown	6 (0.7)	6 (100.0)	–	–	–	–	
Sampling years							0.542[d]
2011	101 (11.3)	91 (90.1)	4 (3.9)	3 (3.0)	1 (1.0)	2 (2.0)	
2012	139 (15.5)	117 (84.2)	11 (8.0)	7 (5.0)	2 (1.4)	2 (1.4)	
2013	102 (11.4)	87 (85.3)	8 (7.8)	4 (3.9)	–	3 (3.0)	
2014	185 (20.6)	160 (86.4)	15 (8.1)	4 (2.2)	2 (1.1)	4 (2.2)	
2015	232 (25.9)	192 (82.8)	18 (7.8)	14 (6.0)	4 (1.7)	4 (1.7)	
2016	137 (15.3)	116 (84.7)	12 (8.8)	6 (4.4)	2 (1.4)	1 (0.7)	

HSX, heterosexuals; MSM, men who have sex with men; IDUs, intravenous drug users; MTCT, mother-to-child transmission.

[a] *HIV subtype A1 (6.2%), C (6.2%), G (18.8%), F1 (6.2%), CRF01_AE (6.2%), CRF03_AB (12.6%), CRF63_02A1 (12.5%), and non-A6/B URFs (31.3%).*

[b] *Value for the difference between HIV-1 subtypes; HIV-1 ≪others≫ subtypes were not considered in the analysis.*

[c] *Pearson's chi-squared test.*

[d] *Pearson's chi-squared test (Yates's correction).*

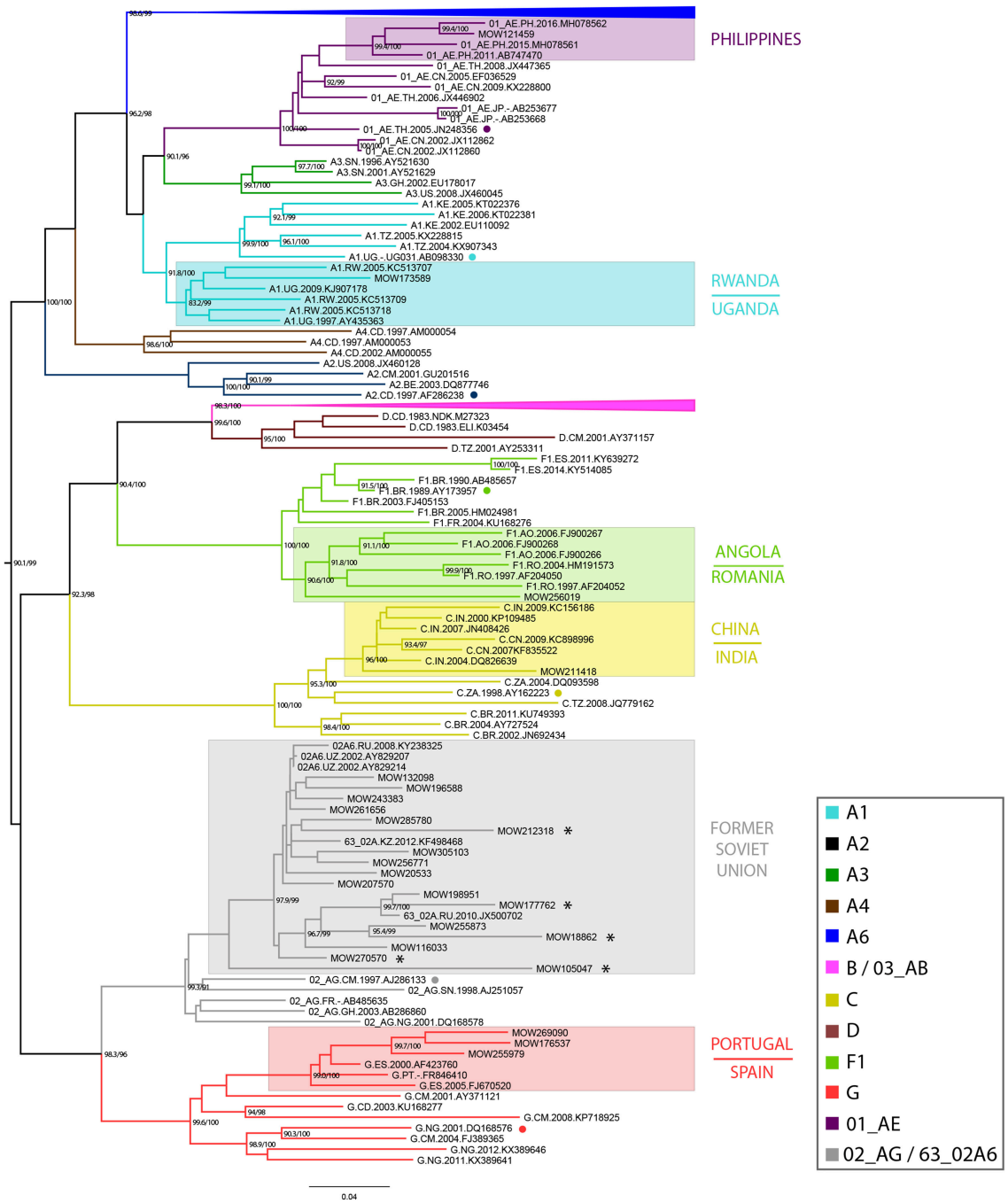


FIGURE 1 | Maximum-likelihood phylogenetic tree of 896 HIV-1 pol sequences from HIV-infected individuals in Moscow region, from 2011 to 2016. The presented tree describes the analysis of all dataset, where the subtype A6 and B clades of HIV-1 are presented in a condensed form. The topology of the A6-clade and B-clade are presented separately (see **Supplemental Figure 1**). Branches are colored according to the HIV-1 subtypes as indicated in the legend. Support values (BS/SH-aLRT) are shown for the key nodes. Asterisks point to locations of unique recombination forms (URFs; see main text). The reference sequence for each subtype was selected using the Reference Alignment from Rega HIV-1 Subtyping Tool and marked by a circle.

Identification of HIV-1 Subtypes A6 and B Epidemic Clusters, and TMRCA Estimation

Maximum-likelihood and Bayesian phylogenetic analyses identified several clusters/subclusters of HIV-1 subtypes A6 and B (**Figure 2** and **Table 2**). All 768 A6-sequences (100%) formed one cluster including sub-clusters (1-2). The proportion of sequences transmitted through HSX, IDU, MSM, and MTCT in these subclusters was comparable. The TMRCA [95% HPD]

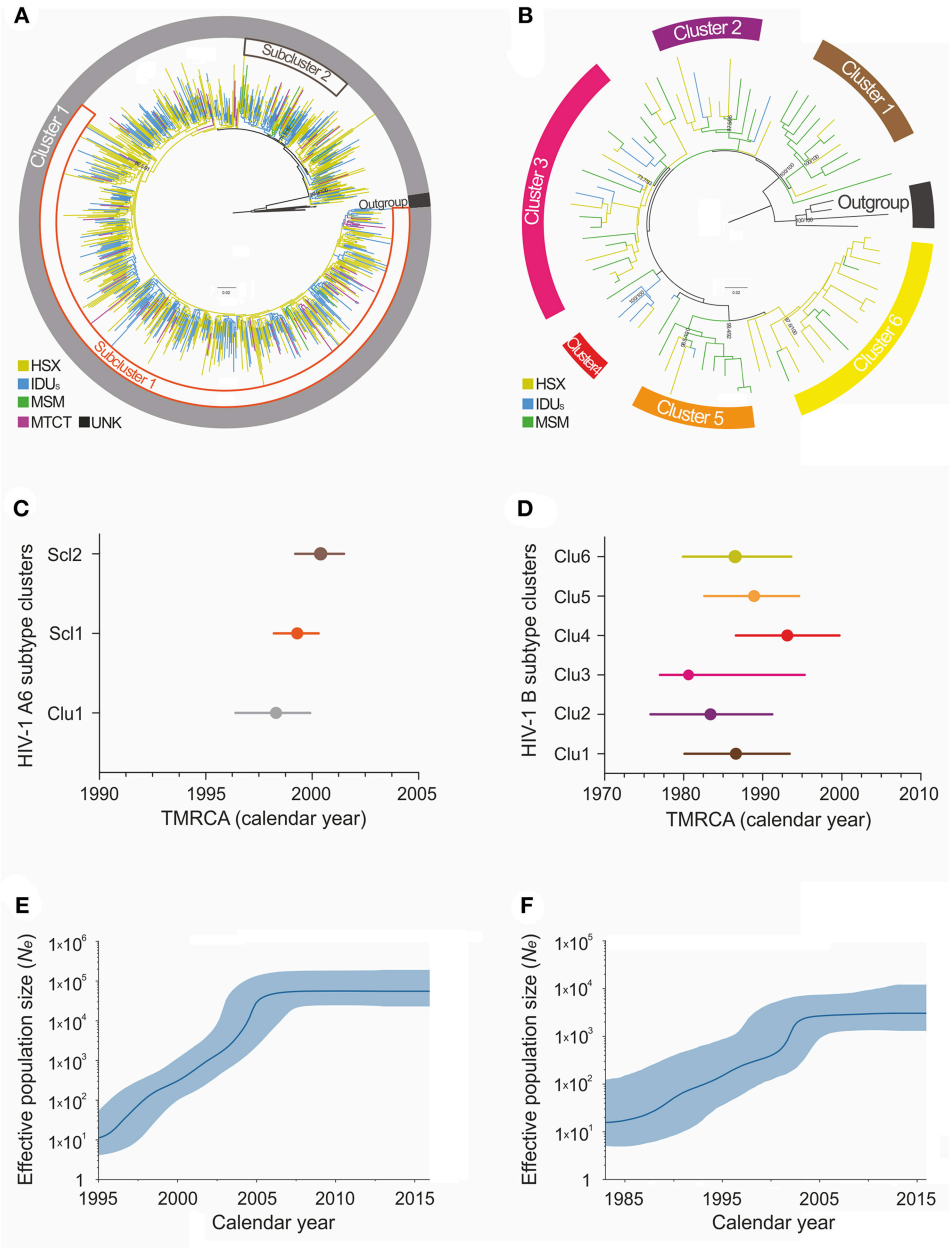


FIGURE 2 | Maximum-likelihood phylogeny and phylodynamic analysis of HIV-1 subtypes in Moscow region. Shown is the analysis of subtypes A6 (left; **A,C,E**) and B (right; **B,D,F**). **(A,B)** ML-tree of the HIV-1 subtypes A and B pol sequences from HIV-infected persons in Moscow, 2007–2016. The colored areas indicate the positions of major epidemic clusters (sub-clusters) identified in the region. Branches are colored according to the risk factor of infection as indicated in the legends. Support values (BS/SH-aLRT) are shown for the key nodes. The trees were rooted through African A1 and D sequences (for A6 and B subtype, respectively). **(C,D)** Bayesian estimates of times of the most recent common ancestors (TMRCA) for the major epidemic clusters (sub-clusters) by BEAST analysis. Points represent the median value, solid lines - 95% credible intervals. **(E,F)** Bayesian skyline plot (BSP) showing effective number of infections over time Ne(t). Estimates of the Ne(t) are presented as median (solid blue line) with the corresponding 95% HPD credibility interval (blue area).

for Cluster 1 and subclusters 1 and 2 were estimated to be 1998.3 [1996.4-1999.9] - 2000.4 [1999.2-2001.5]. Analysis of subtype B sequences identified six clusters (1–6) including 64 (83.2%) sequences. Unlike A6, the genetic structuring of subtype B was to some extent dependent on the transmission route. The smallest cluster (cluster 4) contained ∼5% of the B sequences obtained mainly from IDUs. While the sequences in cluster 3 (largest, ∼25% sequences) were from IDUs and MSM-individuals (in equal proportion), they were predominantly from MSM in clusters 4 and 5, and from HSX-individuals in cluster 2 and 6. The TMRCA for clusters 1-6 were estimated to be 1980.6 [1977.4-1995.3] - 1993.1[1986.6-1999.7] (**Figure 2** and **Table 2**).

TABLE 2 | The cluster sizes, TMRCA, and transmission routes for major epidemic clusters of HIV-1 subtypes in Moscow region.

	Sequence, *N* (%)	Transmission route					*P*-value[a]	TMRCA [95% HPD]
		HSX, *N* (%)	IDUs, *N* (%)	MSM, *N* (%)	MTCT, *N* (%)	Unknown, *N* (%)		
HIV-1 subtype A6							0.136[b]	
Cluster 1	768 (100)	389 (100)	323 (100)	6 (100)	44 (100)	6 (100)		1,998.3 [1,996.4–1,999.9]
Subcluster 1	380 (49.5)	205 (54.0)	144 (37.9)	2 (0.5)	24 (6.3)	5 (1.3)		1,999.3 [1,998.2–2,000.3]
Subcluster 2	90 (11.7)	41 (45.5)	45 (50.0)	0	4 (4.5)	0		2,000.4 [1,999.2–2,001.5]
Un(sub)cluster	298 (38.8)	143 (48.0)	134 (45.0)	4 (1.3)	16 (5.4)	1 (0.3)		–
All	768 (100)	389 (100)	323 (100)	6 (100)	44 (100)	6 (100)		–
HIV-1 subtype B							0.004[c]	
Cluster 1	9 (11.7)	2 (22.2)	0	7 (77.8)	–	–		1,986.6 [1,980.1–1,993.4]
Cluster 2	8 (10.4)	5 (62.5)	1 (12.5)	2 (25.0)	–	–		1,983.4 [1,975.8–1,991.2]
Cluster 3	19 (24.7)	5 (26.3)	4 (21.1)	10 (52.6)	–	–		1,980.6 [1,977.0–1,995.3]
Cluster 4	4 (5.2)	1 (25.0)	3 (75.0)	0	–	–		1,993.1 [1,986.6–1,999.7]
Cluster 5	9 (11.7)	2 (22.2)	1 (11.1)	6 (66.7)	–	–		1,988.9 [1,982.6–1,994.6]
Cluster 6	15 (19.5)	11 (73.4)	0	4 (26.6)	–	–		1,986.5 [1,979.9–1,993.6]
Uncluster	13 (16.8)	5 (38.5)	2 (15.4)	6 (46.1)	–	–		–
All	77 (100)	31 (40.3)	11 (14.3)	35 (45.4)	–	–		–

HSX, Heterosexuals; MSM, men who have sex with men; IDUs, intravenous drug users; MTCT, mother-to-child transmission; TMRCA, time to the most recent common ancestor.
[a] *Value for the difference between clusters (subclusters).*
[b] *MSM and unknown transmission route were not considered in the analysis; Pearson's chi-squared test.*
[c] *Fisher's exact test (two-tailed).*

Demographic History and Evolutionary Rates of HIV-1 Subtypes A6 and B

According to the Bayesian analysis, the mean evolutionary rates [95% HPD] for subtypes A6 and B were 1.26×10^{-3} [1.13×10^{-3}-1.41×10^{-3}] substitutions/site/year and 1.17×10^{-3} [6.10×10^{-4} -1.82×10^{-3}] substitutions/site/year, respectively. The Bayesian Skyline plot indicated at least two effective population growth phases for these subtypes (**Figures 2E,F**). While the population size of A6 grew exponentially between 1998 and 2004, reaching a stable phase (moderate growth) around 2006, subtype B had an initial phase of fast growth between 1985 and 2002 followed by a decline around 2003. Since the mid-2010s, both subtypes have shown a very slow increase.

Recombinant Strains

We found 38 HIV-1 URFs containing genome segments of A6 and B subtypes. Of them, at least 12 (31.6%) had recombinant structures with similar breakpoint locations in the reverse transcriptase region. While the first fragment ($2253\rightarrow$ 3342 ± 65) belonged to sub-subtype A6, the second fragment ($3343 \pm 65\rightarrow$ 3553) belonged to subtype B (**Table 3**).

DISCUSSION

This is the most extensive study in Moscow region to date, devoted to understanding the diversity and temporal dynamics of the common HIV-1 subtypes. We have confirmed that sub-subtype A6 (85.1%) is still the major HIV-1 subtype in the region (Giliazova et al., 2010) (as elsewhere in Russia) (Kazennova et al., 2011; Lapovok et al., 2017), followed by subtype B at 7.6%. The A6 subtype predominates in HSXs and IDUs while subtype B is the most prevalent variant in MSM group. This unequal distribution repeats the molecular pattern of the epidemic in Russia as a whole. Apparently, the ≪ founder effect≫ and the relative social isolation/persistence of individual risk groups (at least at the beginning of the epidemic) are responsible for this phenomenon. We also observed a significant spread (4.2%) of unique recombinants between the two main subtypes, with at least 11 sequences distinct from the currently described A6/B recombinants. We are planning further studies with these recombinants, including full-length genome analysis to determine if they are new circulating forms. The CRF02_AG is the fourth most common HIV variant in the region, which is not surprising given the high level of labor migration from Central Asia where this subtype is widespread. Finally, we have identified at least seven other HIV-1 subtypes, which points to the increasing genetic complexity of the HIV-1 epidemic in the region.

We also investigated the emergence of HIV-1 A6 and B epidemics and their growth rates in Moscow region using the molecular clock. According to our estimates, A6 sub-subtype was introduced into the region around 1998 and marked the onset of massive epidemic (**Figure 2**). These findings are in good agreement with other sero-epidemiological studies (Bobkov et al., 2001). Subsequently, the divergent evolution of A6 strains and possible existence of several independent transmission networks apparently led to the formation of two epidemic sub-clusters, identified in this study.

TABLE 3 | jpHMM-assigned breakpoint locations for 41 HIV-1 unique recombinant sequences in patients from Moscow region.

Sequence name	Fragment (subtype)			Breakpoint†	
	1	2	3	1	2
MOW177762	2,253–3,440 (CRF02)	3,454–3,553 (B)		3,447 ± 6	
MOW212318	2,253–2,322 (B)	2,454–3,553 (CRF02)		2,388 ± 65	
MOW105047	2,253–2,957 (B)	3,029–3,553 (CRF02)		2,993 ± 35	
MOW206415	2,253–2,544 (A6?/B)	2,552–3,352 (A6)	3,404–3,553 (B)	2,548 ± 3	3,378 ± 25
MOW212381	2,253–2,374 (B)	2,450–3,553 (A6)		2,412 ± 37	
MOW191857	2,253–3,397 (A6)	3,421–3,553 (B)		3,409 ± 11	
MOW179103	2,253–3,433 (A6)	3,461–3,553 (B)		3,447 ± 13	
MOW270119	2,253–2,735 (A6)	2,759–3,069 (B)	3,107–3,553 (A6)	2,747 ± 11	3,088 ± 18
MOW26246	2,253–2,312 (B)	2,402–3,553 (A6)		2,357 ± 44	
MOW102645	2,253–2,462 (B)	2,482–3,553 (A6)		2,472 ± 9	
MOW103145	2,253–2,375 (B)	2,517–3,553 (A6)		2,446 ± 70	
MOW109184	2,253–2,374 (B)	2,441–3,553 (A6)		2,408 ± 33	
MOW284149	2,253–2,699 (A6?/A1)	2,733–3,423 (A6)	3,433–3,553 (B)	2,716 ± 16	3,428 ± 4
MOW119218	2,253–3,272 (A6)	3,294–3,553 (B)		3,283 ± 10	
MOW126388	2,253–3,386 (A6)	3,406–3,553 (B)		3,396 ± 9	
MOW125543	2,253–2,474 (A6)	2,528–2,925 (B)	3,061–3,553 (A6)	2,501 ± 26	2,993 ± 67
MOW127873	2,253–3,452 (A6)	3,488–3,553 (B)		3,470 ± 17	
MOW170294	2,253–2,312 (B)	2,358–3,553 (A6)		2,336 ± 22	
MOW173915	2,253–2,397 (B)	2,457–3,553 (A6)		2,427 ± 29	
MOW129062	2,253–3,405 (A6)	3,422–3,491 (B)	3,501–3,553 (A6?/A4)	3,413 ± 8	3,496 ± 4
MOW178169	2,253–3,300 (A6)	3,398–3,553 (B)		3,349 ± 48	
MOW189742	2,253–3,340 (A6)	3,422–3,553 (B)		3,381± 40	
MOW230737	2,253–3,402 (A6)	3,422–3,553 (B)		3,412 ± 9	
MOW236022	2,253–3,295 (A6)	3,333–3,553 (B)		3,314 ± 18	
MOW243257	2,253–3,343 (A6)	3,399–3,553 (B)		3,371 ± 27	
MOW23690	2,253–3,327 (A6)	3,367–3,456 (B)	3,486–3,553 (A6)	3,347 ± 19	3,371 ± 14
MOW257204	2,253–2,362 (B)	2,440–3,553 (A6)		2,401 ± 38	
MOW259399	2,253–3,430 (A6)	3,440–3,553 (B)		3,435 ± 4	
MOW262312	2,253–2,312 (B)	2,374–3,553 (A6)		2,343 ± 30	
MOW263861	2,253–2,321 (B)	2,459–3,553 (A6)		2,390 ± 68	
MOW18862	2,253–2,786 (CRF02)	2,815–2,965 (A6)	3,011–3,553 (CRF02)	2,800 ± 14	2,988 ± 22
MOW269550	2,253–2,312 (B)	2,382–3,553 (A6)		2,347 ± 34	
MOW273948	2,253–3,304 (A6)	3,386–3,553 (B)		3,361 ± 24	
MOW283088	2,253–2,399 (B)	2,457–3,553 (A6)		2,428 ± 28	
MOW285258	2,253–2,666 (B)	2,734–3,553 (A6)		2,700 ± 33	
MOW302647	2,253–2,313 (B)	2,457–3,553 (A6)		2,385 ± 71	
MOW316504	2,253–2,948 (A6)	2,990–3,221 (B)	3,293–3,553 (A6)	2,969 ± 20	3,257 ± 35
MOW328606	2,253–3,342 (A6)	3,398–3,553 (B)		3,370± 27	
MOW243074	2,253–3,392 (A6)	3,422–3,553 (B)		3,407 ± 14	
MOW328087	2,253–3,406 (A6)	3,422–3,553 (A6)		3,414 ± 7	
MOW231027	2,253–3,307 (B)	3,401–3,471 (B)	3,489–3,553 (B)	3,354 ± 46	3,354 ± 8

† *Breakpoint locations (fragment coordinate) correspond to with HXB2 (K03455) numbering. Two HIV-1 sequences (MOW267311 and MOW262544) with multiple recombination points are not represented.*

The study of the genetic relationships between HIV-1 B-strains revealed that the spread of the subtype B in Moscow region involved at least six viral lineages that arose between 1980 and 1993 (**Figure 2**). Clusters 1-3, 5, and 6, which also belong to the Pandemic subtype B were responsible for 70.1% of the sexually transmitted infections and played a dominant role in the B-epidemic in Moscow region among MSM and HSXs. On the other hand, the FSU-B strain (Cluster 4) was responsible for infections in IDUs.

Though we are confident that HIV-1 subtype B penetrated into the region much earlier than A6, the origin of these clusters deserves further investigation. The growth pattern of the B epidemic is similar to that of A6 and represents the combined population dynamics of different subtype B-clusters. With an

initial phase of rapid growth (not as rapid as for subtype A6), it appears that subtype B encountered less favorable conditions for local expansion, such as a slower rate of sexual transmission (Maljkovic Berry et al., 2007). The implementation of ART among all population groups could account for the slowdown in the epidemic since the mid-2010s.

The study can contribute to a better understanding and assessment of the social and biological driving forces behind the development of the epidemic process in one of the key regions for the HIV epidemic in Russia, as well as predicting of future trends of HIV infections and suggesting effective preventive measures.

GENBANK ACCESSION NUMBERS

All HIV-1 sequences described in this study were submitted to Genbank (accession numbers: MH666355-667255; KY857892-857922).

ETHICS STATEMENT

This study was approved by the local Ethics Committee of the Moscow Regional AIDS Centre (RCAC). Informed consent was signed by all the subjects.

AUTHOR CONTRIBUTIONS

AL and MB planned and designed the study. NL and AP collected the epidemiological data and nucleotide sequences. FM performed the recombination analysis and submitted sequences. EK performed the analysis of the epidemiological data. AL performed the phylogenetic and phylodynamic analysis, produced the illustrations and wrote the manuscript. MB supervised the project and edited the manuscript. All authors participated in the critical review of this manuscript.

FUNDING

The research was funded by the Russian Science Foundation, agreement N 15-15-00050-P (2018-2019). The collecting of samples was supported from the sources of Ministry of education and science grant (number RFMEFI61019X0020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00320/full#supplementary-material

REFERENCES

Bobkov, A., Kazennova, E., Khanina, T., Bobkova, M., Selimova, L., Kravchenko, A., et al. (2001). An HIV type 1 subtype A strain of low genetic diversity continues to spread among injecting drug users in Russia: study of the new local outbreaks in Moscow and Irkutsk. *AIDS Res. Hum. Retroviruses* 17, 257–261. doi: 10.1089/088922201750063188

Bobkov, A., Kazennova, E., Selimova, L., Bobkova, M., Khanina, T., Ladnaya, N., et al. (1998). A sudden epidemic of HIV type 1 among injecting drug users in the former Soviet Union: identification of subtype A, subtype B, and novel gagA/envB recombinants. *AIDS Res. Hum. Retroviruses* 14, 669–676. doi: 10.1089/aid.1998.14.669

Bobkova, M. (2013). Current status of HIV-1 diversity and drug resistance monitoring in the former USSR. *AIDS Rev.* 15, 204–212.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Diez-Fuertes, F., Cabello, M., and 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

Dukhovlinova, E., Masharsky, A., Toussova, O., Verevochkin, S., Solovyeva, T., Meringof, M., et al. (2015). Two independent HIV epidemics in Saint Petersburg, Russia revealed by molecular epidemiology. *AIDS Res. Hum. Retroviruses* 31, 608–614. doi: 10.1089/aid.2014.0150

FedAC. (2010). *Report on the Russian HIV/AIDS Epidemic*. Available online at: http://www.hivrussia.ru/files/bul_34.pdf

Foley, B. T., Leitner, T., Paraskevis, D., and Peeters, M. (2016). Primate immunodeficiency virus classification and nomenclature: review. *Infect. Genet. Evol.* 46, 150–158. doi: 10.1016/j.meegid.2016.10.018

Gashnikova, N. M., Bogachev, V. V., Baryshev, P. B., Totmenin, A. V., Gashnikova, M. P., Kazachinskaya, A. G., et al. (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

Gashnikova, N. M., Safronov, P. F., Nikonorova, Y. V., Unagaeva, N. V., Lapteva, T. A., Bogachev, V. V., et al. (2011). Properties of CRF02_AG HIV-1 isolates circulating in Novosibirsk region. *Zh. Mikrobiol. Epidemiol. Immunobiol.* 38–43.

Gashnikova, N. M., Zyryanova, D. P., Astakhova, E. M., Ivlev, V. V., Gashnikova, M. P., Moskaleva, N. V., et al. (2017). Predominance of CRF63_02A1 and multiple patterns of unique recombinant forms of CRF63_A1 among individuals with newly diagnosed HIV-1 infection in Kemerovo Oblast, Russia. *Arch. Virol.* 162, 379–390. doi: 10.1007/s00705-016-3120-4

Giliazova, A. V., Zenin, P. V., Pronin, A., Serkov, I. L., Khametova, K. M., Orlova-Morozova, E. A., et al. (2010). Molecular epidemiology of HIV-1 strains in the Moscow Region. *Vopr. Virusol.* 55, 25–29.

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 25, 679–689. doi: 10.1097/QAD.0b013e328342ff93

Johnson, V. A., Calvez, V., Gunthard, H. F., Paredes, R., Pillay, D., Shafer, R., et al. (2011). 2011 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 19, 156–164.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* doi: 10.1093/bib/bbx108. [Epub ahead of print].

Kazennova, E., Laga, V., Gromov, K., Lebedeva, N., Zhukova, E., Pronin, A., et al. (2017). Genetic variants of HIV type 1 in men who have sex with men in Russia. *AIDS Res. Hum. Retroviruses* 33, 1061–1064. doi: 10.1089/aid.2017.0078

Kazennova, E. V., Antonova, O. V., Kuzin, S. N., Serkina, T. P., Sokolova, L. S., Vasil'ev, A. V., et al. (2011). Molecular and epidemiology studies of HIV-1 prevalence in the Republic of Sakha (Yakutia). *Vopr Virusol* 56, 30–34.

Kostaki, E. G., Karamitros, T., Bobkova, M., Oikonomopoulou, M., Magiorkinis, G., Garcia, F., et al. (2018). Spatiotemporal characteristics of the HIV-1 CRF02_AG/CRF63_02A1 epidemic in Russia and Central Asia. *AIDS Res. Hum. Retroviruses* 34, 415–420. doi: 10.1089/aid.2017.0233

Lapovok, I., Kazennova, E., Laga, V., Vasilyev, A., Utegenova, A., Abishev, A., et al. (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

Lapovok, I. A., Lopatukhin, A. E., Kireev, D. E., Kazennova, E. V., Lebedev, A. V., Bobkova, M. R., et al. (2017). Molecular epidemiological analysis of

HIV-1 variants circulating in Russia in 1987-2015. *Ter. Arkh.* 89, 44–49. doi: 10.17116/terarkh2017891144-49

Maljkovic Berry, I., Ribeiro, R., Kothari, M., Athreya, G., Daniels, M., Lee, H. Y., et al. (2007). Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J. Virol.* 81, 10625–10635. doi: 10.1128/JVI.00985-07

Nabatov, A. A., Kravchenko, O. N., Lyulchuk, M. G., Shcherbinskaya, A. M., and Lukashov, V. V. (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. doi: 10.1089/08892220260190380

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Pineda-Pena, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., et al. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* 19, 337–348. doi: 10.1016/j.meegid.2013.04.032

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Schultz, A. K., Bulla, I., Abdou-Chekaraou, M., Gordien, E., Morgenstern, B., Zoaulim, F., et al. (2012). jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus. *Nucleic Acids Res.* 40, W193–198. doi: 10.1093/nar/gks414

Struck, D., Lawyer, G., Ternes, A. M., Schmit, J. C., and 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., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4:vey016. doi: 10.1093/ve/vey016

Thomson, M. M., Vinogradova, A., Delgado, E., Rakhmanova, A., Yakovlev, A., Cuevas, M. T., 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. doi: 10.1097/QAI.0b013e31819c1757

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer DP declared a past co-authorship with one of the authors MB to the handling editor.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 March 2019
doi: 10.3389/fmicb.2019.00369

A 28-Year History of HIV-1 Drug Resistance and Transmission in Washington, DC

Keylie M. Gibson[1], Margaret C. Steiner[1], Seble Kassaye[2], Frank Maldarelli[3], Zehava Grossman[3,4], Marcos Pérez-Losada[1,5,6] and Keith A. Crandall[1,6]*

[1] Computational Biology Institute, Milken Institute School of Public Health, George Washington University, Washington, DC, United States, [2] Department of Medicine, Georgetown University, Washington, DC, United States, [3] HIV Dynamics and Replication Program, Host-Virus Interaction Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States, [4] Sackler Faculty of Medicine, School of Public Health, Tel Aviv University, Tel Aviv, Israel, [5] CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal, [6] Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, George Washington University, Washington, DC, United States

Washington, DC consistently has one of the highest annual rates of new HIV-1 diagnoses in the United States over the last 10 years. To guide intervention and prevention strategies to combat DC HIV infection, it is helpful to understand HIV transmission dynamics in a historical context. Toward this aim, we conducted a retrospective study (years 1987–2015) of 3,349 HIV *pol* sequences (1,026 bp) from 1,995 individuals living in the DC area belonging to three different cohorts. We coupled HIV sequence data with clinical information (sex, risk factor, race/ethnicity, viral load, subtype, anti-retroviral regimen) to identify circulating drug resistant mutations (DRM) and transmission clusters and assess their persistence over time. Of the transmission clusters identified in the DC area, 78.0 and 31.7% involved MSM and heterosexuals, respectively. The longest spread of time for a single cluster was 5 years (2007–2012) using a distance-based network inference approach and 27 years (1987–2014) using a maximum likelihood phylogenetic approach. We found eight subtypes and nine recombinants. Genetic diversity increased steadily over time with a slight peak in 2009 and remained constant thereafter until 2015. Nucleotide diversity also increased over time while relative genetic diversity (BEAST) remained relatively steady over the last 28 years with slight increases since 2000 in subtypes B and C. Sequences from individuals on drug therapy contained the highest total number of DRMs (1,104–1,600) and unique DRMs (63–97) and the highest proportion (>20%) of resistant individuals. Heterosexuals (43.94%), MSM (40.13%), and unknown (44.26%) risk factors showed similar prevalence of DRMs, while injection drug users had a lower prevalence (33.33%). Finally, there was a 60% spike in the number of codons with DRMs between 2007 and 2010. Past patterns of HIV transmission and DRM accumulation over time described here will help to predict future efficacy of ART drugs based on DRMs persisting over time and identify risk groups of interest for prevention and intervention efforts within the DC population. Our results show how longitudinal data can help to understand the temporal dynamics of HIV-1 at the local level.

Keywords: Washington, DC, phylodynamics, HIV-1, drug resistance mutations, transmission networks

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Teiichiro Shiino,
National Institute of Infectious Diseases (NIID), Japan
Bin Su,
Beijing Youan Hospital, Capital Medical University, China

****Correspondence:***
Keylie M. Gibson
kmgibson@gwu.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *03 December 2018*
Accepted: *12 February 2019*
Published: *08 March 2019*

Citation:
Gibson KM, Steiner MC, Kassaye S, Maldarelli F, Grossman Z, Pérez-Losada M and Crandall KA (2019) A 28-Year History of HIV-1 Drug Resistance and Transmission in Washington, DC. Front. Microbiol. 10:369. doi: 10.3389/fmicb.2019.00369

INTRODUCTION

The prevention and treatment of HIV in Washington, DC is a public health priority both locally and nationally, as more diagnoses per unit population occur annually in the District than in any of the 50 states (Centers for Disease Control and Prevention, 2017). According to the most recent report by the DC Department of Health (DOH), over 1,900 per 100,000 residents are living with HIV in DC as of 2017 (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2018). This surpasses the WHO's threshold for a generalized epidemic (1%) (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017). Additionally, the epidemic disproportionately affects black men and women, who comprise 75% of HIV cases but only 46.7% of residents in DC (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017). Seven of the eight wards in DC have an HIV prevalence greater than 1%, indicating that the epidemic is not entirely geographically localized (Pérez-Losada et al., 2017). Similar to the Chicago HIV epidemic (Morgan et al., 2017), it has been proposed that the severity of the DC epidemic may be due to the high proportion of residents who belong to high risk groups and the city's considerable economic stratification (Greenberg et al., 2009).

The United States HIV epidemic is composed of predominantly subtype B variants (Vermund and Leigh-Brown, 2012), although several studies report that non-B subtypes are becoming more prominent over time both in DC (Grossman et al., 2018), the US (Germer et al., 2015) and globally (Ivanov et al., 2013; Neogi et al., 2014; UK Collaborative Group on HIV Drug Resistance, 2014; Esbjörnsson et al., 2016). Notably, of the individuals living with HIV in DC, more than 10% were born outside of the United States, thus leading to a potentially high diversity of HIV in the region (Grossman et al., 2018). Additionally, with a variety of subtypes and recombinant forms circulating in Washington, DC, the presence of drug resistant mutations (DRM) becomes a relevant clinical issue that needs to be addressed in order to provide efficient and effective treatment. Drug resistance is one of the most pressing issues in HIV treatment today, as resistance to preventative and treatment drugs affects the ability of the virus to evade the immune system and persist in the infected individual. Drug resistant mutations may occur at the time of transmission, so treatment-naïve individuals may also be resistant at the time of diagnosis. Currently, there are 97 known codons affected by DRMs in the Stanford HIV database (https://hivdb.stanford.edu), all either conferring resistance to or under surveillance (SDRM) for one or more of the current 22 antiretroviral (ART) drugs. Resistance to antiretroviral drugs and ensuing viremia can result in immunosuppression leading to morbidity, and, if new DRMs arise, then new drugs are needed to combat the ever-evolving virus. In 2016, Kassaye et al. found a high prevalence of drug resistance mutations in persons living with HIV in DC. Thus, studying the presence and spread of DRMs in Washington, DC can provide a temporal and spatial context to study the dynamics of HIV drug resistance. Furthermore, identifying DRMs present and spreading between risk groups will provide public health officials with information to implement more efficient targeted education and prevention efforts and programs.

According to the DC DOH, men who have sex with men (MSM) and heterosexuals (HRH) are the primary modes of HIV transmission in DC, while transmission via injection drug use (IDU) has decreased by 95% from 2007 to 2017 (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017). About 27% of people living with HIV are black MSM, while the second and third most prominent HIV risk groups are heterosexual black women and white MSM (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017). A cross-sectional study in 2009 identified heterosexual transmission as an emerging epidemic in non-Hispanic black men and women in DC, demonstrating that this risk is based on transmission networks rather than individual risk (Magnus et al., 2009). Similarly, a study of transmission clusters in Sweden, Denmark, and Finland suggested MSM-heterosexual transmission played a significant role in HIV infection since the majority of the mixed-transmission clusters involved individuals from those two risk groups (Esbjörnsson et al., 2016).

To address the current HIV epidemic in Washington, DC—and to better recognize how it will evolve in the future—we must first understand the evolutionary dynamics of the local epidemic. Toward this aim, we conducted a retrospective study (years 1987–2015) of 1,995 HIV-infected individuals whose 3,349 sequences were collected from three separate datasets. Our analysis focused specifically on the polymerase (*pol)* gene, the most frequently sequenced gene in HIV-1 clinical studies of DRMs. We identified transmission clusters with associations to demographic information and mapped epidemiological variables on estimated transmission clusters. Lastly, we identified drug resistance mutations and assessed their variation over time, as well as estimated sequence diversity over time.

MATERIALS AND METHODS

Datasets

A total of 3,349 HIV sequences (after removing duplicates) from the metropolitan DC area (including Washington, DC, northern Virginia, and northern and southern Maryland) were combined from three independently collected datasets: Pérez-Losada et al. (2017) with 1,659 sequences one each from that same number of individuals, Maldarelli et al. (2013) with 1,387 sequences collected from 33 individuals, and new HIV data presented here representing 303 sequences, one per individual (**Table 1**). All duplicate sequences in Pérez-Losada et al.'s dataset were removed, but all sequences in Maldarelli et al.'s dataset were used as it included known intra-patient variation and reflected the evolution of the virus over time. All studies performed DNA Sanger sequencing. Collectively, we have a total of 1,995 patients with sequence samples obtained between 1987 and 2015. Demographic variables considered were: sex, gender, race/ethnicity, age, country of birth, state of residence, risk factor,

TABLE 1 | Phenotypic characteristics of each dataset separated and combined.

	(Maldarelli et al., 2013)	(Pérez-Losada et al., 2017)	This study	Total
Number of individuals	33	1,659	303	1,995
Number of sequences	1,387	1,659	303	3,349
Median age (years)	35.5	46	39	44.79
Age range (years)	19.5, 51.5	10, 82	20.2, 75	10, 82
Quartiles (years)	28.5, 41.9	34, 54	31.8, 46.5	30, 50
RACE/ETHNICITY				
Black	NA	1,394 (84.0%)	NA	1,394 (69.9%)
White	NA	138 (8.3%)	NA	138 (6.9%)
Hispanic	NA	76 (4.6%)	NA	76 (3.8%)
Other	NA	22 (1.3%)	NA	22 (1.1%)
Unknown	33 (100%)	29 (1.8%)	303 (100%)	365 (18.3%)
SEX AT BIRTH				
Male	NA	1,117 (67.3%)	NA	1,117 (56.0%)
Female	NA	542 (32.7%)	NA	542 (27.2%)
Unknown	33 (100%)	0 (0%)	303 (100%)	336 (16.8%)
GENDER				
Male	31 (93.9%)	1,117 (67.3%)	229 (75.6%)	1,377 (69.0%)
Female	2 (6.1%)	543 (32.7%)	72 (23.8%)	617 (30.9%)
Transgender	0 (0%)	0 (0%)	2 (0.7%)	2 (0.1%)
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)
COUNTRY OF BIRTH				
US	NA	114 (6.9%)	NA	114 (5.7%)
Non-US	NA	50 (3.01%)	NA	50 (2.5%)
Unknown	33 (100%)	1,495 (90.1%)	303 (100%)	1,831 (91.8%)
STATE OF RESIDENCE				
DC	NA	1,368 (82.5%)	303 (100%)	1,671 (83.8%)
MD	NA	127 (7.7%)	0 (0%)	127 (6.4%)
VA	NA	28 (1.7%)	0 (0%)	28 (1.4%)
Other	NA	0 (0%)	0 (0%)	0 (0%)
Unknown	33 (100%)	136 (8.2%)	0 (0%)	169 (8.5%)
HIV RISK FACTOR				
MSM	0 (0%)	637 (38.4%)	138 (45.5%)	775 (38.8%)
IDU	0 (0%)	86 (5.2%)	6 (1.98%)	92 (4.61%)
Heterosexual	0 (0%)	558 (33.6%)	152 (50.2%)	710 (35.6%)
Sex	33 (100%)	0 (0%)	2 (0.7%)	35 (1.75%)
Other	0 (0%)	378 (22.8%)	3 (1.0%)	381 (19.1%)
Unknown	0 (0%)	303 (18.3%)	2 (0.7%)	305 (15.3%)
Median CD4 count (cells/μl)	401	498	274.5	419
VIRAL LOAD (COPIES/ML)				
<200	NA	129 (7.8%)	39 (12.9%)	168 (%)
200–399	NA	32 (1.9%)	4 (1.3%)	36 (%)
400–9,999	NA	245 (14.8%)	63 (20.8%)	308 (%%)
>10,000	NA	371 (22.4%)	154 (50.8%)	525 (%)
Unknown	33 (100%)	882 (53.2%)	43 (14.2%)	958 (48.0%)
ART EXPOSURE				
Experienced	0 (0%)	1,005 (60.6%)	0 (0%)	1,005 (50.4%)
Naïve	0 (0%)	106 (6.4%)	303 (100%)	409 (20.4%)
Unknown	33 (100%)	548 (33.0%)	0 (0%)	581 (29.1%)
ART REGIMEN TYPE				
Single Class				

(Continued)

TABLE 1 | Continued

	(Maldarelli et al., 2013)	(Pérez-Losada et al., 2017)	This study	Total
NRTI	NA	5 (0.3%)	NA	5 (0.3%)
NNRTI	NA	2 (0.1%)	NA	2 (0.1%)
PI	NA	6 (0.4%)	NA	6 (0.3%)
ENH	NA	4 (0.2%)	NA	4 (0.2%)
INT	NA	7 (0.4%)	NA	7 (0.4%)
Dual Class	NA	59 (3.6%)	NA	59 (3.0%)
Multiple Class	NA	1,455 (87.7%)	NA	1,455 (72.9%)
Unknown	33 (100%)	121 (7.3%)	NA	154 (7.7%)
DRM Present	438 (31.6%)	647 (39.0%)	185 (61.1%)	1,270 (37.9%)
SUBTYPE				
B	33 (100%)	1,559 (93.9%)	225 (74.3%)	1,817 (91.1%)
C	0 (0%)	32 (1.9%)	16 (5.3%)	48 (2.4%)
D	0 (0%)	7 (0.4%)	4 (1.3%)	11 (0.6%)
B recombinant	0 (0%)	38 (2.3%)	21 (6.9%)	59 (2.9%)
Other	0 (0%)	23 (1.4%)	37 (12.2%)	60 (3.0%)
SEQUENCING DATE				
1987–1994	1 (3.0%)	0 (0%)	0 (0%)	1 (0.05%)
1995–2000	14 (42.4%)	0 (0%)	7 (2.3%)	21 (1.1%)
2001–2005	18 (54.5%)	0 (0%)	45 (14.9%)	63 (3.2%)
2006–2010	0 (0%)	0 (0%)	112 (37.0%)	112 (5.6%)
2011–2015	0 (0%)	1,659 (100%)	139 (45.9%)	1,798 (90.1%)

DC, District of Columbia; MD, Maryland; VA, Virginia; MSM, men who have sex with men; IDU, injection drug users; ART, antiretroviral therapy; DRM, drug resistance mutations; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors; PI, protease inhibitors; ENH, enhancers; INT, integrase inhibitors.

viral load, duration of infection, CD4+T lymphocyte count, HIV-1 subtype, and antiretroviral regimen type. These characteristics were paired with their respective HIV-1 sequence(s) (**Table 1**).

Briefly, sequences in the Pérez-Losada et al. (2017) dataset were collected through the DC Cohort, a longitudinal study conducted by the DC Department of Health at 13 DC area clinics (Milken Institute School of Public Health, 2018). Over 1,700 patients were enrolled in this study between 2011 and 2015. RNA from plasma samples was sequenced by LabCorp as part of routine clinical care with a focus on the *pol* genes. RT-PCR and Sanger sequencing were used to generate reads, which were analyzed with Sequencher DNA Sequence Analysis Software. Sequences for *PR/RT* sequences had a length of 1,496 bp. This dataset included additional integrase sequences (864 bp), which were not included in our analysis.

The dataset originally published in Maldarelli et al. (2013) was comprised of 33 patients treated at the NIH Clinical Center in Bethesda, Maryland. HIV sequences were obtained via single genome Sanger sequencing from plasma samples; limiting dilution was completed on each plasma sample. These patients were sampled longitudinally. Amplicons covering protease and reverse transcriptase (297 and 700–1200 bp, respectively) were used to obtain *PR/RT* sequences.

Data collection and sequencing of new HIV data was completed with the same single genome sequencing procedure as in Maldarelli et al. (2013). However, longitudinal sampling of these patients was not completed. A single plasma sample was taken from each patient at time of sampling and only a single limiting dilution was completed. These patients were also from the NIH Clinical Center in Bethesda, MD.

Phylodynamic Analyses

We aligned all sequences to the HXB2 reference *pol* sequence using MAFFT (Katoh et al., 2002). Aligned sequences were trimmed to a 1,026 bp region (2,253–3,279 bp relative to HXB2; GenBank accession K03455), which covered all of *PR* and a region of *RT* (55%). Any nucleotides outside this gene region were trimmed to ensure consistency among datasets. We used jModelTest (Posada, 2008) to estimate the best-fit model of molecular evolution along with model parameter values used in phylogenetic inference (Posada and Crandall, 2001). We constructed a maximum-likelihood (ML) phylogenetic tree using RAxML v. 8.2.9 (Stamatakis, 2014).

We assessed transmission clusters using both phylogenetic analyses and the network method HIV-TRACE (Kosakovsky Pond et al., 2018). HIV-TRACE identifies transmission clusters by analyzing genetic distance between pairs of sequences. For HIV-TRACE, we used a genetic distance threshold of 0.01 substitutions/site (Wertheim et al., 2017; as in Pérez-Losada et al., 2017) to account for the genetic diversity of *pol*. Additionally, we handled the ambiguities in the sequences using an average to account for all possible resolutions and avoid biases and false positive transmission clusters between the different datasets and sequencing strategies. Default settings were used for the remaining parameters. For the ML phylogenetic method, a

TABLE 2 | Nucleotide and genetic diversity estimates.

	Subtype B					B Recombinants					Subtype C				
Subset	***N***	***h***	***S***	**π**	**θ (W)**	***N***	***h***	***S***	**π**	**θ (W)**	***N***	***h***	***S***	**π**	**θ (W)**
All	3085	50	1005	0.05323	0.07934	156	103	522	0.0656	0.08882	48	44	362	0.05582	0.07875
DATASETS															
Frank	1387	1045	851	0.04975	0.07219	0	NA	NA	NA	NA	0	NA	NA	NA	NA
NIH	222	54	497	0.05587	0.08414	34	34	351	0.0647	0.0883	16	16	208	0.0533	0.06561
LC	1476	60	656	0.05612	0.0812	122	114	494	0.06612	0.09006	32	32	314	0.05691	0.07693
GENDER															
Male	2503	60	983	0.05215	0.0788	112	105	493	0.06581	0.08914	26	26	287	0.05554	0.07287
Female	580	43	607	0.0506	0.07563	43	43	379	0.0644	0.08674	22	22	258	0.05576	0.06902
Transgender	1	NA	NA	NA	NA	1	NA	NA	NA	NA	0	NA	NA	NA	NA
Unknown	1	NA	NA	NA	NA	0	NA	NA	NA	NA	0	NA	NA	NA	NA
RACE															
Black	1244	65	644	0.05569	0.08212	101	96	477	0.06596	0.0902	25	25	277	0.05559	0.07258
Hispanic	64	64	428	0.0577	0.08879	9	9	198	0.06333	0.07169	2	2	63	0.06294	0.06294
White	128	116	483	0.05622	0.08723	7	7	176	0.06499	0.07083	0	NA	NA	NA	NA
Other	17	17	302	0.06634	0.08759	1	NA	NA	NA	NA	3	3	95	0.06348	0.06217
Unknown	1632	140	653	0.04904	0.08001	38	38	363	0.06503	0.08703	18	18	224	0.05731	0.06854
RISK FACTOR															
HRH	572	39	596	0.05568	0.08175	60	60	416	0.06327	0.08594	32	31	309	0.05493	0.07448
IDU	89	78	429	0.04635	0.08106	3	3	89	0.06022	0.05957	0	NA	NA	NA	NA
MSM	690	74	616	0.05669	0.08216	76	75	452	0.06571	0.08832	4	4	100	0.05518	0.05421
MSM/IDU	17	17	266	0.05718	0.07743	1	NA	NA	NA	NA	0	NA	NA	NA	NA
SEX	1387	1045	851	0.04975	0.07219	0	NA	NA	NA	NA	0	NA	NA	NA	NA
Other	57	56	435	0.06141	0.09184	3	3	125	0.08447	0.08447	0	NA	NA	NA	NA
Unknown	273	75	529	0.05409	0.08398	13	13	241	0.06618	0.07677	12	12	192	0.0559	0.06344
DATE															
1987	38	30	77	0.01307	0.01666	0	NA	NA	NA	NA	0	NA	NA	NA	NA
1990	11	10	60	0.01472	0.01864	0	NA	NA	NA	NA	0	NA	NA	NA	NA
1996	22	21	66	0.01284	0.01692	0	NA	NA	NA	NA	0	NA	NA	NA	NA
1997	49	47	183	0.03227	0.0375	0	NA	NA	NA	NA	0	NA	NA	NA	NA
1998	19	16	114	0.01698	0.02754	0	NA	NA	NA	NA	0	NA	NA	NA	NA
1999	117	109	306	0.03949	0.0393	0	NA	NA	NA	NA	0	NA	NA	NA	NA
2000	350	263	479	0.03788	0.05151	1	NA	NA	NA	NA	0	NA	NA	NA	NA
2001	220	173	400	0.03225	0.04716	0	NA	NA	NA	NA	0	NA	NA	NA	NA
2002	254	192	483	0.04924	0.06361	0	NA	NA	NA	NA	0	NA	NA	NA	NA
2003	158	130	328	0.04532	0.0538	2	2	48	0.05263	0.05263	0	NA	NA	NA	NA
2004	183	126	354	0.04267	0.05251	0	NA	NA	NA	NA	1	NA	NA	NA	NA
2005	8	8	172	0.06214	0.07291	0	NA	NA	NA	NA	1	NA	NA	NA	NA
2006	6	6	145	0.0659	0.06982	1	NA	NA	NA	NA	0	NA	NA	NA	NA
2007	15	15	191	0.04953	0.06561	4	4	74	0.04702	0.04529	0	NA	NA	NA	NA
2008	15	15	236	0.05976	0.07967	2	2	49	0.05438	0.05438	1	NA	NA	NA	NA
2009	23	23	277	0.05836	0.0776	2	2	67	0.07306	0.07306	2	2	41	0.06376	0.06386
2010	23	23	281	0.05671	0.07723	8	8	180	0.06538	0.0712	2	2	31	0.04641	0.04641
2011	580	65	607	0.05528	0.08359	48	48	408	0.06444	0.08874	18	18	223	0.05199	0.06321
2012	429	77	588	0.05623	0.08437	41	40	401	0.0674	0.09022	14	14	223	0.05729	0.06786
2013	292	59	543	0.05605	0.0827	29	29	340	0.06419	0.08257	6	6	134	0.06069	0.05992
2014	152	126	501	0.05882	0.08671	10	10	223	0.06691	0.07791	2	2	66	0.06483	0.06483
2015	121	112	465	0.05682	0.08485	8	8	197	0.06855	0.07456	1	NA	NA	NA	NA

N, number of sequences; h, number of haplotypes; S, number of polymorphic sites; π, nucleotide diversity; θ (W), Watterson's genetic diversity.

transmission cluster was defined as a node with ≥70 bootstrap support (Felsenstein, 1985).

Watterson's genetic diversity (θ), nucleotide diversity (π), haplotype diversity (h), and the number of polymorphic sites (S) were estimated separately for HIV subtypes B and C and recombinant B and patient subsets (e.g., risk types, sex, race/ethnicity, date, and dataset) in DnaSP v. 6.11.01 (Rozas et al., 2017). Relative genetic diversity over time was inferred separately for subtypes B and C and recombinant B in BEAST2 (Bouckaert et al., 2014) using the GMRF Bayesian Skyride model (Minin et al., 2008) and the date of the HIV-1 sample as calibration and a normal prior with mean = 0.001 and *SD* = 0.0005 for ucld.mean. Additionally, we used the HKY substitution model with gamma-distributed among-site rate heterogeneity, a relaxed clock (log-normal) model of rate of substitution and partitioned into three codon positions. We performed two runs 5×10^8 generations each with sampling every 5,000 generations. Parameter uncertainty was summarized in the 95% highest posterior density (HPD) intervals. Results generated by BEAST were visualized in Tracer (Rambaut et al., 2018) with a 10% burn-in.

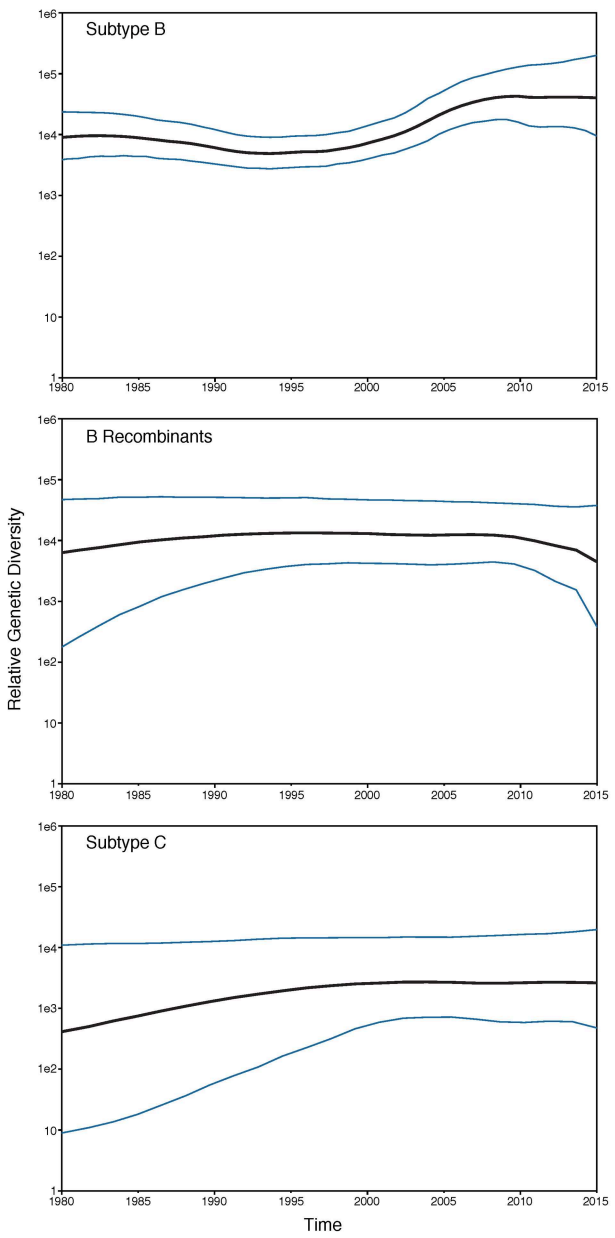


FIGURE 1 | GMRF Bayesian Skyride plot of HIV-1 subtypes past population dynamics. The 95% high posterior density limits of the relative genetic diversity over time are shown in blue. Black lines show the median estimate of the relative genetic diversity over time.

Drug Resistance Mutations and Subtype Analyses

We identified drug resistance mutations using the HIVdb program (Liu and Shafer, 2006) from the Stanford University HIV Drug Resistance Database (https://hivdb.stanford.edu). HIVdb identifies known resistance and surveillance variants (SDRMs) associated with 22 approved antiretroviral therapies including Protease Inhibitors (PI), Nucleoside RT Inhibitors (NRTI), Non-Nucleoside RT Inhibitors (NNRTI), and Integrase Inhibitors (INSTI), as well as treatment-selected mutations (TSMs). In order to compare DRMs over time, the number of affected codons for sequences from a single year was normalized by the number of sequences from that year in the dataset. We identified HIV-1 subtypes and recombinants phylogenetically using 170 subtype and recombinant reference sequences from the Los Alamos HIV database (http://www.hiv.lanl.gov/) and ML tree reconstruction, as well as using the REGA subtyping tool v. 3.0 provided by Stanford University (Pineda-Peña et al., 2013). Although subtyping was already reported for the Pérez-Losada et al. dataset, we included these sequences again for confirmation and comparison to the additional datasets from Washington, DC.

RESULTS

The number of sequences included in our analyses varied between 0 and 540 sequences per year, with seven years missing between 1987 and 2015. The cohort was comprised primarily of men who have sex with men (MSM; 38.8%) and high-risk heterosexuals (HRH; 35.6%) (**Table 1**). The median age at the time of sequencing was 44.8 years, with lower and upper bounds of 10.0 and 82.0 years old, respectively. Antiretroviral exposure information was only available for Pérez-Losada et al.'s dataset, and the majority of those patients were on multiple-class drug regimens (57.93%). REGA and phylogenetic subtyping methods identified eight unique HIV-1 subtypes and nine recombinants in our combined dataset, of which 10 were neither B or B recombinants (BD, CRF12_BF, DB, CRF24_BG, BF, and BA). Those subtypes were B (94.95%), C (1.43%), and A, D, G, F1, F2, J, (<1%) and those recombinants were BD (1.37%), and CRF02_AG, CRF12_BF, DB, CRF18_cpx, CRF24_BG, BF, BA, and AK (<1%).

Overall, Washington, DC was found to have a high genetic diversity (mean Watterson's θ = 0.082, **Table 2**), with

B recombinants having a slightly higher genetic diversity ($\theta = 0.089$) compared to subtype B ($\theta = 0.079$) and subtype C ($\theta = 0.079$) (**Table 2**). For all subtypes, males had a higher genetic diversity than females, and for subtype B, all races/ethnicities had similar genetic diversity. However, for B recombinants and subtype C, blacks had a higher genetic diversity than the remaining races/ethnicities. For subtype B, there was an overall high genetic diversity for all risk groups, with general sexual contact (SEX) risk having the lowest genetic diversity ($\theta = 0.079$). This is not unexpected given that all 1,387 HIV sequences from patients with general sexual contact risk belong to Maldarelli et al.'s dataset and are only from 33 patients. Furthermore, subtype B injection drug users (IDU) had the lowest nucleotide diversity ($\pi = 0.046$), while MSM and HRH had similar nucleotide diversity estimates. Moreover, genetic diversity in subtype B increased steadily over time, with a slight decrease in diversity in 2003 and 2004 (**Table 2**). Nucleotide diversity increased over time with a peak in 2005 and 2006. No steady trends of diversity were estimated in B recombinants and subtype C. Additionally, the past demographic analyses of subtype B, B recombinants, and subtype C in BEAST (**Figure 1**) indicated that HIV relative genetic diversity remained stable over the last 35 years, with subtype B showing more variability.

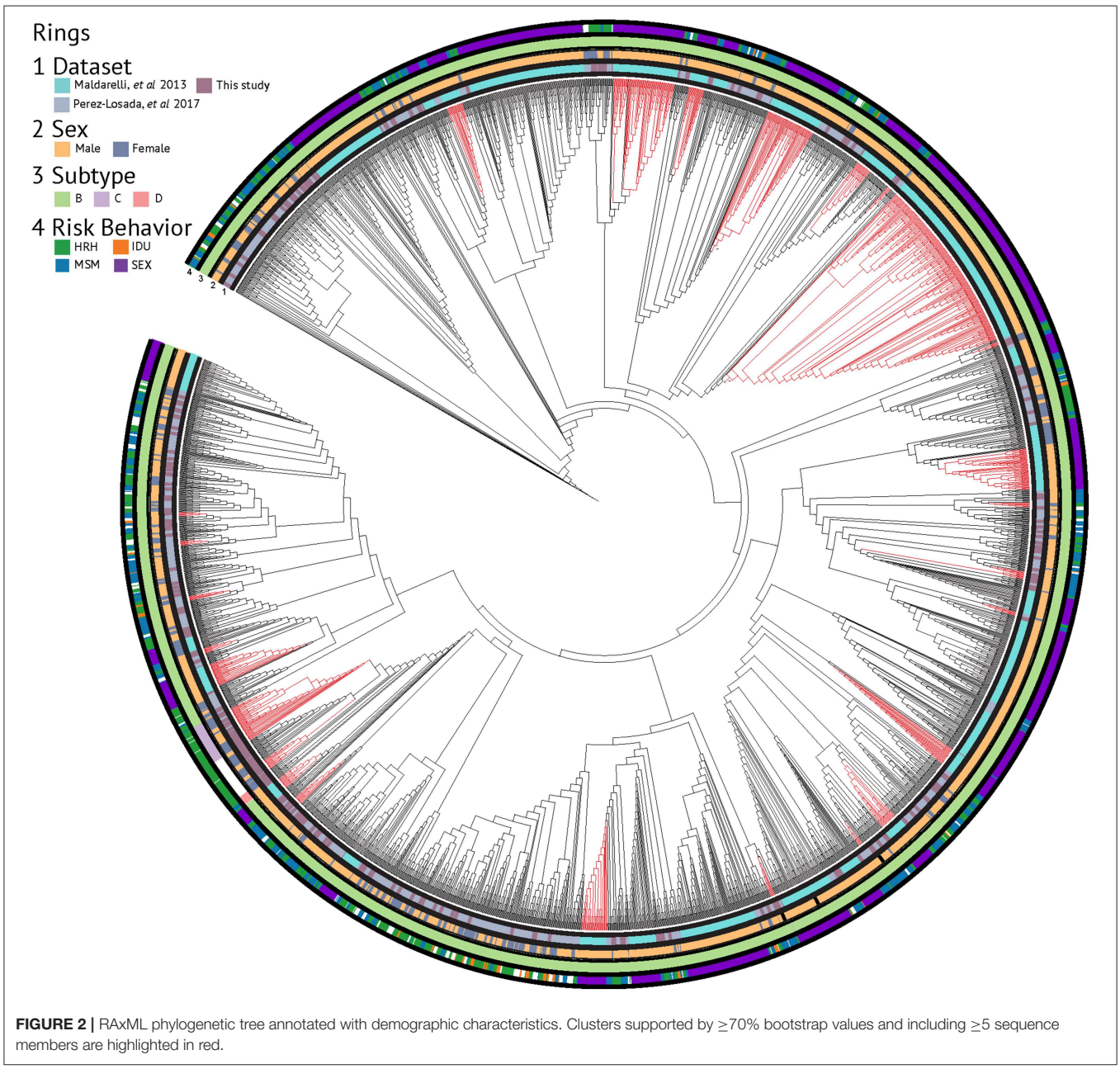


FIGURE 2 | RAxML phylogenetic tree annotated with demographic characteristics. Clusters supported by ≥70% bootstrap values and including ≥5 sequence members are highlighted in red.

Evolutionary relationships amongst the HIV sequences were represented by a star phylogeny, indicating equal dispersal around the DC area (**Figure 2**). We annotated the phylogenetic tree with four demographic characteristics (dataset, sex, subtype, and risk behavior) and found near even dispersal for all of them. Sequences from each of the 33 Maldarelli et al.'s patients clustered together both in RAxML with high support and in HIV-TRACE, as expected; however, some of the intra-patient sequences clustered into more than one cluster. Clusters that were comprised of sequences from a single Maldarelli et al.'s patient were not included in the total cluster number count (RAxML = 27, HIV-TRACE = 30). A total of 1,798 sequences (53.7%) were grouped into 215 clusters in the RAxML tree, with 16 of them belonging to groups with >5 individuals (**Figure 2**). Of the clusters that were comprised of Maldarelli et al.'s patients, 12 clusters contained at least one sequence from a different dataset (most often the new data). MSM comprised 59.7 and 85.7% of all clusters and clusters with ≥5 individuals, respectively (**Table 3**). A total of 40 clusters were comprised of MSM and HRH samples, while only 8 contained an IDU sample (**Table 3**). HIV-TRACE transmission networks were labeled by combinations of three phenotypic characteristics: race, risk factor, and sex (**Figure 3**, **Table 3**). A total of 314 (9.4%) sequences were grouped into 41 clusters comprised of two or three individuals. Of these clusters, 78.0% included at least one MSM individual and 31.7% included at least one HRH individual. A third (27.9–31.7%) of the clusters predicted from either method included ≥2 individuals from different datasets. The longest spread of time for a single cluster found in HIV-TRACE and RAxML was 5 years (2007–2012) and 27 years (1987–2014), respectively. The majority of sequences in clusters found for HIV-TRACE methods were from the same year (61%), unlike RAxML, where only a quarter (27.4%) of the sequences in the transmission clusters were from the same year. Overall, RAxML included more individuals in more transmission clusters compared to those predicted by HIV-TRACE.

TABLE 3 | Phenotypic characteristics in transmission clusters.

Trait	HIV-TRACE clusters containing ≥1 individual (% of total)	RAxML phylogenetic clusters containing ≥1 individual (% of total)
Total clusters	**41**	**215**
SEX		
2 individuals	37 (90.2)	144 (67.0)
3 individuals	4 (9.8)	44 (20.5)
4 individuals	0 (0.0)	11 (5.1)
5+ individuals	0 (0.0)	16 (7.4)
RISK GROUP		
HRH	13 (31.7)	120 (55.8)
IDU	1 (2.4)	8 (3.7)
M&I	1 (2.4)	7 (3.3)
MSM	32 (78.0)	129 (60.0)
SEX	7 (17.1)	15 (7.0)
Other	3 (7.3)	14 (6.5)
Unknown	5 (12.2)	67 (31.2)
SEX		
Female	10 (24.4)	108 (50.2)
Male	37 (78.7)	179 (83.3)
DATASET		
(Maldarelli et al., 2013)	7 (17.1)	15 (7.0)
(Pérez-Losada et al., 2017)	32 (78.0)	195 (90.7)
This study	16 (39.0)	68 (31.6)
RACE		
Black	27 (65.9)	177 (82.3)
Hispanic	6 (14.6)	18 (8.4)
Other	1 (2.4)	8 (3.7)
Unknown	12 (29.3)	73 (34.0)
White	3 (7.3)	26 (12.1)

HRH, high risk heterosexual; IDU, injection drug user; MSM, men who have sex with men; M&I, men who have sex with men and injection drug user; SEX, sexual activity (general).

Drug resistance mutations (DRMs) were analyzed for subtype B sequences only (95% of the total sequence data). Across the combined dataset comprising all three studies, 24 amino acids in *PR* and 61 in *RT* were affected by drug resistance mutations. The types of antiretroviral drugs with the highest DRM prevalence were nucleoside reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), and reverse transcriptase surveillance drug resistance mutations (RT SDRMs); these drugs were found to have between 64 and 97 unique DRMs present in our dataset (**Figure 4**, **Table 4**), with unique DRMs referring to a specific DRM observed at a codon position in one or more sequences. Therefore, if a mutation was detected more than once it was not double counted. Additionally, more than 20% of the individuals in the combined dataset displayed at least one DRM for one or more of these antiretroviral (ART) drugs. The risk group with the highest prevalence of DRMs was blood transfusion/perinatal (58.33%), but this is likely due to the low number of blood transfusion/perinatal individuals in our dataset (44 individuals, 1.3%). Additionally, these individuals often begin ART regimens almost immediately upon birth or acquisition of HIV, thus providing a longer time for DRMs to develop and persist in their viral population. Heterosexuals and MSM were found to have similar DRM prevalence at 43.94 and 40.13%, respectively. Those individuals with unknown risk factors had a DRM prevalence of 44.26%, and a lower prevalence was seen in injection drug users at 33.33% (**Figure 4**). Notably, a 60% increase in the number of codons affected by DRMs occurred between 2006 and 2010 (**Figure 5**). The amount of DRMs for PR Major (DRMs that make a major contribution to reduced susceptibility to protease inhibitors), PR Accessory (DRMs that contribute to reduced susceptibility in combination with PR Major DRMs), PR SDRMs, NRTI, NNRTI, RT SDRMs, and protease inhibitor treatment-selected mutations (PI TSMs) increased notably in the more recent years (2011–2015); however, this is likely due to the high number of sequences included in those years (**Table 4**). NRTI and RT SDRMs consistently had higher number of sequences that contained ≥1 DRM, total mutations, and unique mutations from 1996 onward.

Fewer unique mutations were found to be at higher frequencies in the 1990s, whereas more unique mutations were found to be at a lower frequency in the 2000s (**Table 4**, **Figure 6**). This trend can be explained by the number of sequences per year,

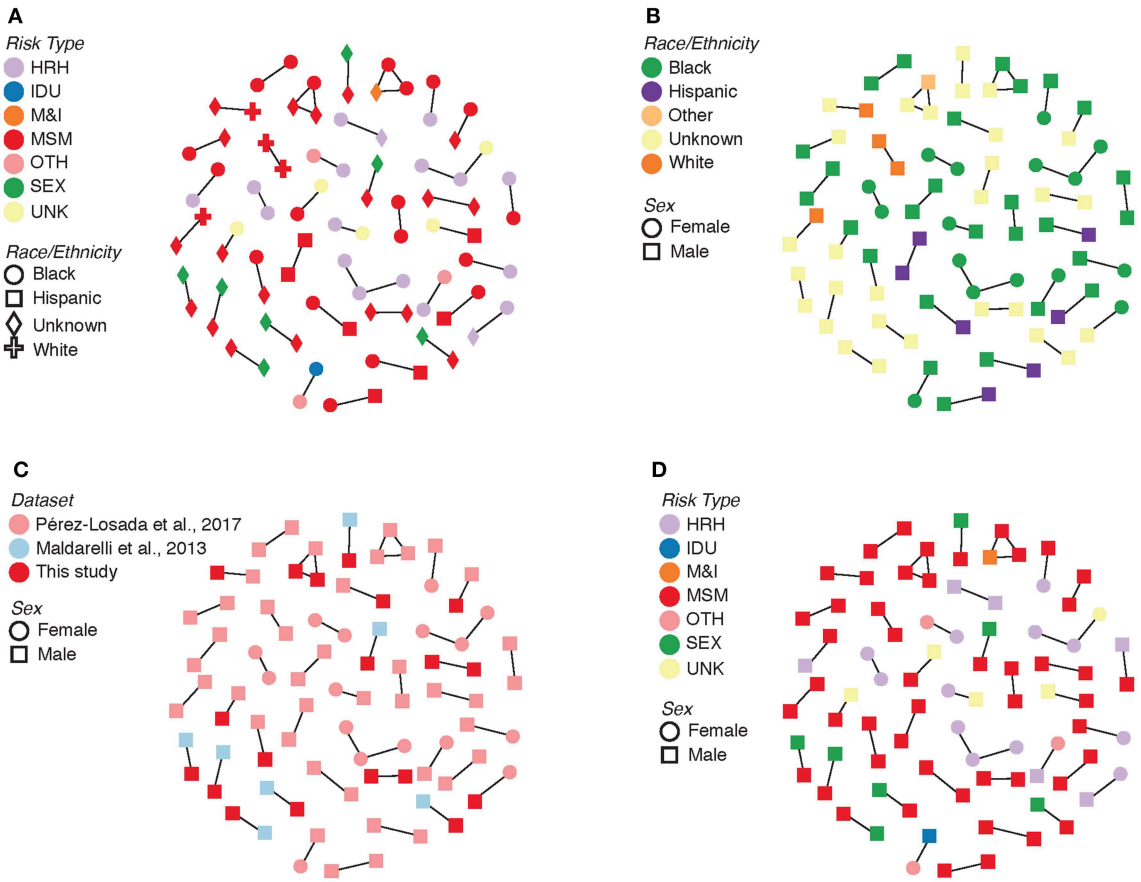


FIGURE 3 | Transmission clusters with associated demographic information. **(A)** Risk type by race/ethnicity. **(B)** Race/ethnicity by sex. **(C)** Dataset by sex. **(D)** Risk type by sex. HRH, high risk heterosexual; IDU, injection drug user; MSM, man who has sex with men; SEX, sexual activity (general); MSM&IDU, MSM and IDU; OTH, other; UNK, unknown.

where there were lower total sequences in the 1990s and early 2000s (**Table 4**). However, despite there being a lower number of total sequences between 2005 and 2010, there were more distinct mutations in that time range (**Figure 6**) as well as a peak in the number of affected codons (**Figure 5**), as seen in a recent study of treatment-naïve individuals in the US (Ross et al., 2018). Notably, in 1996 there were eight DRMs found at high frequencies (*PR*: D30N, N88D; *RT*: D67N, K70R, Y181C, M184V, K219E, H22IY); all eight DRMs belonged to a single patient from Maldarelli et al.'s dataset (patient 27), in which one blood sample from a single appointment was obtained and 22 serial dilutions resulted in 22 separate sequences (Maldarelli et al., 2013). Two mutations in *RT*, K103N and M184V, persisted from around 2005 to 2015 at a lower frequency (average 12.4 and 8.8%, respectively). K103N is a Major NNRTI resistant mutation, whereas M184V is a Major NRTI resistant mutation.

DISCUSSION

Transmission Networks

Multiple studies have identified transmission networks as playing a significant role in the spread of HIV in the United States and DC (Magnus et al., 2009; Pérez-Losada et al., 2010, 2017; Esbjörnsson et al., 2016; Kassaye et al., 2016; Morgan et al., 2017; Wertheim et al., 2017). We identified in our dataset 41 and 76 unique transmission clusters using HIV-TRACE and a ML phylogenetic method (RAxML), respectively, most of which included members of the MSM and HRH risk groups. Transmission within these high-risk groups is expected given that both groups are reported as continuing sources of HIV infection in Washington, DC (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017, 2018). A majority of clusters included at least one non-Hispanic black individual, which is consistent with public health surveillance data as well (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2018). Like Kassaye et al. (2016), the majority of our identified transmission clusters were comprised of only two or three patients (81.9–100%). The ML and HIV-TRACE results included several transmission clusters containing individuals from different datasets (27.9 and 31.7%, respectively), suggesting that many of those networks are unique to this study. By combining sequencing and demographic information across a 28-year time window, we found that individuals in transmission networks predicted by HIV-TRACE had sequences obtained from different

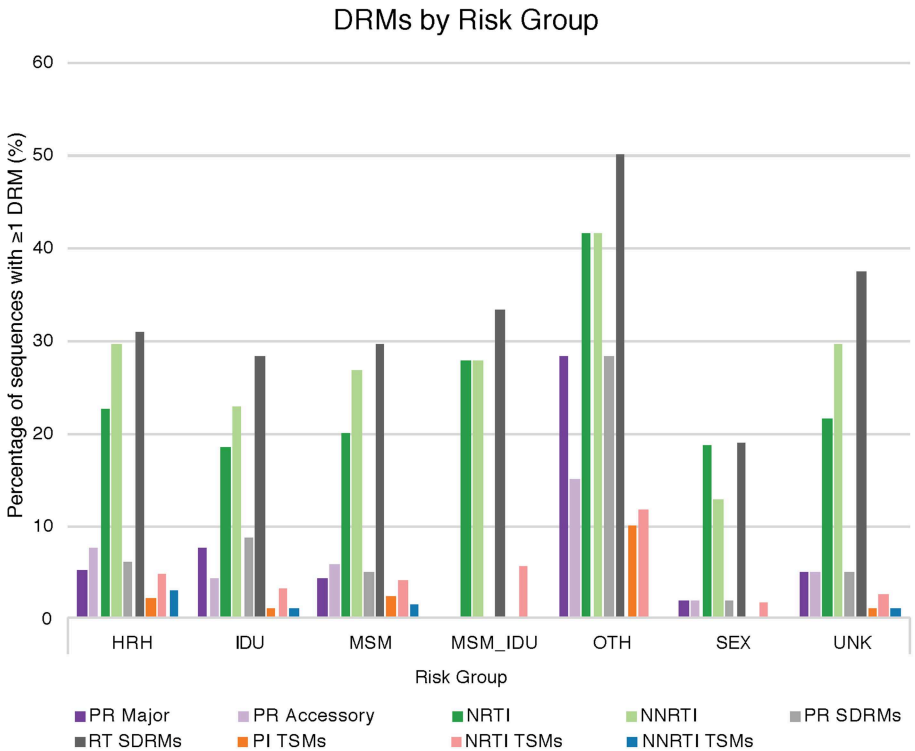


FIGURE 4 | Drug resistance mutations by risk group. The percentage of sequences with at least one drug resistance mutation in a given ART drug category provided by Stanford HIVdb, separated by HIV risk group. HRH, high risk heterosexual; IDU, injection drug user; MSM, man who has sex with men; SEX, sexual activity (general); MSM_IDU, MSM and IDU; OTH, other; UNK, unknown; PR Major, DRMs that make a major contribution to reduced susceptibility to protease inhibitors; PR Accessory, DRMs that contribute to reduced susceptibility in combination with PR Major DRMs; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors; PR SDRMs, protease surveillance drug resistant mutations; RT SDRMs, reverse transcriptase surveillance drug resistant mutations; PI TSMs, protease inhibitor treatment-selected mutations; NRTI TSMs, nucleoside reverse transcriptase inhibitors treatment-selected mutations; NNRTI TSMs, non-nucleoside reverse transcriptase inhibitors treatment-selected mutations.

years ranging between 1 and 5 years apart, with a mean of 2.5 year spread in sequence time. However, the majority of the clusters for HIV-TRACE contained sequences from the same year (61.0%). For the RAxML, on the other hand, the longest amount of time between sequences in a transmission cluster was 27 years; although there was a mean of 2.4 year spread in sequence time. These results indicate that viral sequences do not often persist over many years, but when they do, distance-based cluster analyses might have a more difficult time identifying such transmission networks compared to evolutionary phylogenetic approaches. Our results show how insights from longitudinal studies can be used to understand viral dynamics over time in a local epidemic as well as how methodological choices can impact final inferences in such studies.

Temporal Diversity of HIV in DC

HIV diversity, as indicated by the number of unique subtypes and recombinants, and genetic diversity increased in the latter years (most recent sampling) of our study. Subtype B (94.95%) and B recombinants (2.9%) dominated our HIV dataset, followed by subtype C (1.43%). As suggested by Grossman et al. (2018), subtype C has been in DC since the late 1980s; however, our sampling of those same years did not include subtype C sequences. Similarly to a recent study of the San Francisco area over a 14-year time period, we also found over 4% non-subtype B HIV-1 variants in our combined dataset (Dalai et al., 2018). Focusing on subtype B, genetic diversity was similar across race/ethnicity but was higher in HRH and MSM relative to IDU risk factors. Genetic diversity increased steadily over time with a slight peak in 2009 and remained constant thereafter until 2015 ($\theta = \sim 0.086$). Additionally, π, which is indicative of current genetic diversity (Crandall et al., 1999), increased over time as well, further suggesting that variants are accumulating in the DC population. The number of haplotypes generally increased until 2005 followed by a sharp reduction of haplotypes but a consistent genetic diversity. This trend could be a consequence of the small sample sizes in years 2005–2010, but relatively low numbers of haplotypes remained in years 2011–2013 despite high sample sizes. Our phylodynamic analyses of the three dominant subtypes in our dataset (B, B recombinants, and C) showed that the relative genetic diversity of DC's HIV population has remained relatively steady over the last 35 years, with subtype C showing an increase in relative genetic diversity and subtype B showing an increase in diversity since 2000. These results coupled with the phylogenetic tree being star-like further bolsters the conclusion that the HIV

TABLE 4 | Drug resistant mutations over time.

		1987			1990			1996			1997		
Number of sequences (total)		**38**			**11**			**22**			**49**		
		N	**TM**	**UM**	**N**	**TM**	**UM**	**N**	**TM**	**UM**	**N**	**TM**	**UM**
Protease	PR Major	0	0	0	0	0	0	22	22	1	0	0	0
	PR Accessory	0	0	0	0	0	0	22	22	1	1	1	1
	PR SDRMs	0	0	0	0	0	0	22	44	2	0	0	0
Reverse Transcriptase	NRTI	0	0	0	0	0	0	22	86	7	0	0	0
	NNRTI	0	0	0	1	1	1	22	40	5	0	0	0
	RT SDRMs	0	0	0	1	1	1	22	111	10	0	0	0
	PI TSMs	0	0	0	0	0	0	1	1	1	0	0	0
	NRTI TSMs	0	0	0	0	0	0	1	1	1	0	0	0
	NNRTI TSMs	0	0	0	0	0	0	0	0	0	0	0	0
		1998			**1999**			**2000**			**2001**		
Number of sequences (total)		**19**			**117**			**350**			**220**		
Protease	PR Major	1	1	1	0	0	0	3	3	3	3	3	3
	PR Accessory	1	1	1	0	0	0	1	1	1	0	0	0
	PR SDRMs	1	1	1	0	0	0	3	3	3	3	3	3
Reverse Transcriptase	NRTI	17	18	3	19	20	4	139	139	1	13	13	3
	NNRTI	1	1	1	21	21	1	4	4	3	2	2	2
	RT SDRMs	17	19	4	17	17	1	142	142	3	13	13	3
	PI TSMs	0	0	0	0	0	0	0	0	0	0	0	0
	NRTI TSMs	0	0	0	0	0	0	1	1	1	0	0	0
	NNRTI TSMs	0	0	0	0	0	0	1	1	1	1	1	1
		2002			**2003**			**2004**			**2005**		
Number of sequences (total)		**254**			**158**			**183**			**8**		
Protease	PR Major	0	0	0	0	0	0	1	1	1	0	0	0
	PR Accessory	4	4	4	0	0	0	2	2	2	2	2	2
	PR SDRMs	0	0	0	0	0	0	1	1	1	0	0	0
Reverse Transcriptase	NRTI	29	31	7	2	3	2	24	24	2	4	6	5
	NNRTI	7	7	6	54	54	4	76	80	9	1	1	1
	RT SDRMs	30	33	8	5	6	3	24	28	6	3	5	4
	PI TSMs	0	0	0	0	0	0	2	2	2	0	0	0
	NRTI TSMs	1	1	1	1	1	1	20	20	1	1	1	1
	NNRTI TSMs	1	1	1	0	0	0	0	0	0	0	0	0
		2006			**2007**			**2008**			**2009**		
Number of sequences (total)		**5**			**14**			**15**			**23**		
Protease	PR Major	1	4	4	0	0	0	1	4	4	1	1	1
	PR Accessory	2	2	2	3	3	3	4	6	6	3	3	3
	PR SDRMs	2	5	5	0	0	0	3	7	6	1	2	2
Reverse Transcriptase	NRTI	2	11	9	3	6	6	5	9	7	5	10	8
	NNRTI	2	2	2	4	5	3	6	9	7	7	10	6
	RT SDRMs	3	12	10	5	10	8	6	12	8	7	16	10
	PI TSMs	0	0	0	0	0	0	0	0	0	0	0	0
	NRTI TSMs	1	2	2	1	1	1	0	0	0	3	3	3
	NNRTI TSMs	1	1	1	1	1	1	0	0	0	2	2	2
		2010			**2011**			**2012**			**2013**		
Number of sequences (total)		**23**			**585**			**431**			**295**		
Protease	PR Major	0	0	0	33	63	22	25	43	20	16	37	12
	PR Accessory	3	3	2	42	61	20	24	35	16	14	19	11

(Continued)

TABLE 4 | Continued

		2010			2011			2012			2013		
Number of sequences (total)		**38**			**11**			**22**			**49**		
		N	**TM**	**UM**	**N**	**TM**	**UM**	**N**	**TM**	**UM**	**N**	**TM**	**UM**
	PR SDRMs	0	0	0	34	79	23	31	57	20	17	37	13
Reverse Transcriptase	NRTI	9	10	7	131	226	64	89	171	57	65	131	44
	NNRTI	11	15	9	166	258	55	126	191	51	85	138	40
	RT SDRMs	11	17	10	191	377	51	140	276	48	90	212	42
	PI TSMs	0	0	0	16	16	7	10	10	6	8	9	4
	NRTI TSMs	0	0	0	25	27	9	16	16	6	17	21	7
	NNRTI TSMs	2	2	1	6	7	6	7	8	5	4	4	2
		2014			**2015**								
Number of sequences (total)		**151**			**121**								
Protease	PR Major	10	20	16	5	6	3						
	PR Accessory	7	10	9	1	1	1						
	PR SDRMs	10	22	13	5	7	4						
Reverse Transcriptase	NRTI	26	49	29	21	32	16						
	NNRTI	35	51	23	30	38	15						
	RT SDRMs	46	81	30	31	50	17						
	PI TSMs	3	3	2	0	0	0						
	NRTI TSMs	4	5	3	2	3	3						
	NNRTI TSMs	0	0	0	4	4	4						

N, number of sequences with ≥1 DRM; TM, total number of mutations; UM, number of unique mutations; PR Major, DRMs that make a major contribution to reduced susceptibility to protease inhibitors; PR Accessory, DRMs that contribute to reduced susceptibility in combination with PR Major DRMs; PR SDRMs, protease surveillance drug resistant mutations; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors; RT SDRMs, reverse transcriptase surveillance drug resistant mutations; PI TSMs, protease inhibitor treatment-selected mutations; NRTI TSMs, nucleoside reverse transcriptase inhibitors treatment-selected mutations; NNRTI TSMs, non-nucleoside reverse transcriptase inhibitors treatment-selected mutations.

epidemic in DC is mature and individuals with varying infection durations and risk factor have been intermingling for years (Kassaye et al., 2016).

Evolution of Drug Resistant Mutations in DC

Our study found a higher percentage of patients with DRMs (37.9%) than a previous study (Kassaye et al., 2016) of the DC epidemic (22.5%). This is expected since our study also included a higher proportion of treatment-experienced individuals, while Kassaye et al. involved only ART-naïve patients. The DRM classes designated by the Stanford University HIV Drug Resistance Database (NRTI, NNRTI, and RT SDRMs) were notably more prominent in our data. There was little variation between the number of DRM-affected codons between risk groups.

We found an increase in the number of codons affected by DRMs between 2005 and 2008. It is possible that the smaller number of sequences from the years 2005 through 2010 relative to other years may have contributed to an artificial inflation of the number of DRMs during that time period. Additionally, all sequences from 2005 through 2010 came from the new dataset and were all treatment-naïve. A similar trend and spike in DRM prevalence was found in treatment-naïve patients in the U.S. between 2000 and 2009 (Ross et al., 2018), which correlates with the increase in diversity in subtype B. Such a trend may have been caused by changes in ART recommendations that allowed treatment to begin more quickly after diagnosis, the approval of additional therapies, mainly integrase inhibitors and specifically Maraviroc (MVC) and Raltegravir (RAL) introduced in 2007 (see **Figure 5**), and the increased use of genotypic screening for drug resistance prior to treatment (Ross et al., 2018). Kassaye et al. (2016) also observed a downward trend in DRMs over time between 1997–2006 and 2007–2013. Furthermore, medication guidelines recommended by physicians to patients have changed over time, which could result in the variation of number of DRMs and number of codon positions affected by mutation. Moreover, for the treatment-experienced individuals, HIV sequencing was likely requested because of concerns about their adherence and presumably because they are still viremic while being on medication, contributing to the increase in these numbers as well.

Although there was a decline in the number of DRM-affected codons around 2008, the number of unique drug resistant mutations increased throughout the 2000s. All of these DRMs were found at low frequencies, but a similar upward trend was also seen in genetic and nucleotide diversity. With increased diversity and increased DRMs, combating HIV replication and spread becomes more cumbersome, thus requiring researchers and doctors to stay one step ahead of the epidemic by developing new drugs that are not already impaired by the current DRMs before the current drugs are no longer effective. This is partially corrected already by patients in dual- or multiple-class ART regimens. Two mutations, K103N and M184V both in *RT*, persisted for 10 years in this combined dataset. K103N is

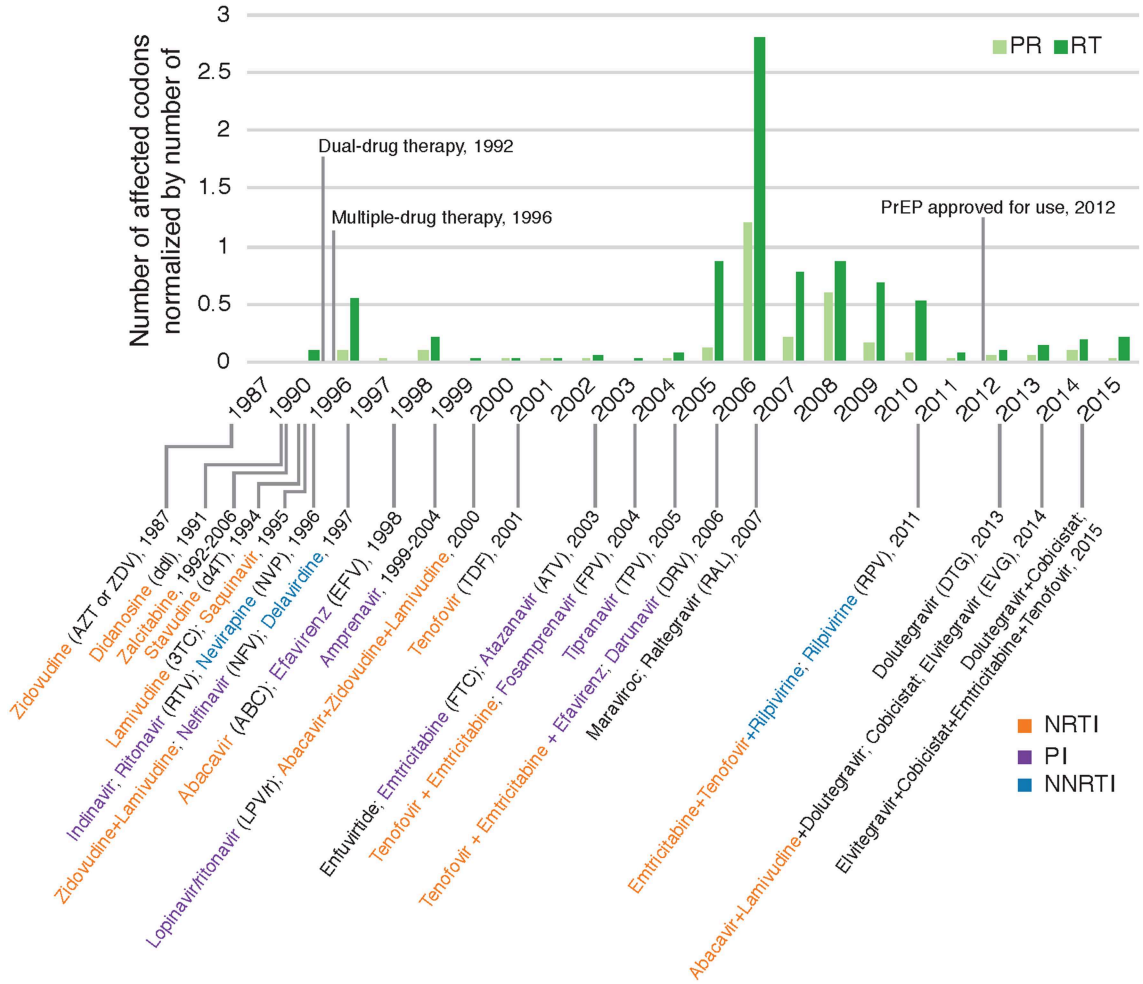


FIGURE 5 | Drug resistant mutations annotated with relevant ART dates. Count of DRM-affected codons observed in each year normalized by the total number of sequences collected in that year. Dates with relevant drug therapy and introductions were included (Arts et al., 2012; Barré-Sinoussi et al., 2013; U.S. Food Drug Administration, 2018). NRTI, nucleoside reverse transcriptase inhibitors; PI, protease inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors.

selected for by usage of nevirapine (NVP) and efavirenz (EFV) (Bacheler et al., 2000; Gulick et al., 2004; Margot et al., 2006); M184V is selected by lamivudine (3TC) and emtricitabine (FTC) (Keulen et al., 1996; Frost et al., 2000). These four drugs are commonly prescribed as part of treatment recommendations (Geneva: World Health Organization, 2016, 2017).

Limitations

This study has a few limitations. Sampling across years is inconsistent and heterogeneous with more recent years containing more individuals and HIV sequences than older years. Additionally, initial years of the DC HIV epidemic (1980s) are missing. This uneven sampling could shorten the duration of the observed transmission clusters. This is further compounded when using HIV-TRACE, because a genetic distance cut-off is used, limiting the transmission clusters predicted by HIV-TRACE to more similar sequences. We supplemented the transmission clusters predicted by HIV-TRACE with transmission clusters predicted by phylogenetic methods, where genetic distance is not a cut-off (only branch support is), thus allowing for more distantly related viruses to be included in predicted transmission clusters. Each dataset was sequenced in different laboratories, leading to potential biases in ambiguity codes and consensus sequence resolution that could affect diversity estimates and potentially produce false positive or negative transmission clusters. Moreover, the direction of HIV virus transmission in a transmission cluster was not determined.

CONCLUSIONS

Like several other studies in the US (Latkin et al., 2011; DiNenno et al., 2012; Raymond et al., 2017) and in DC (Kassaye et al., 2016), our study supports the DC DOH's observation that HRH

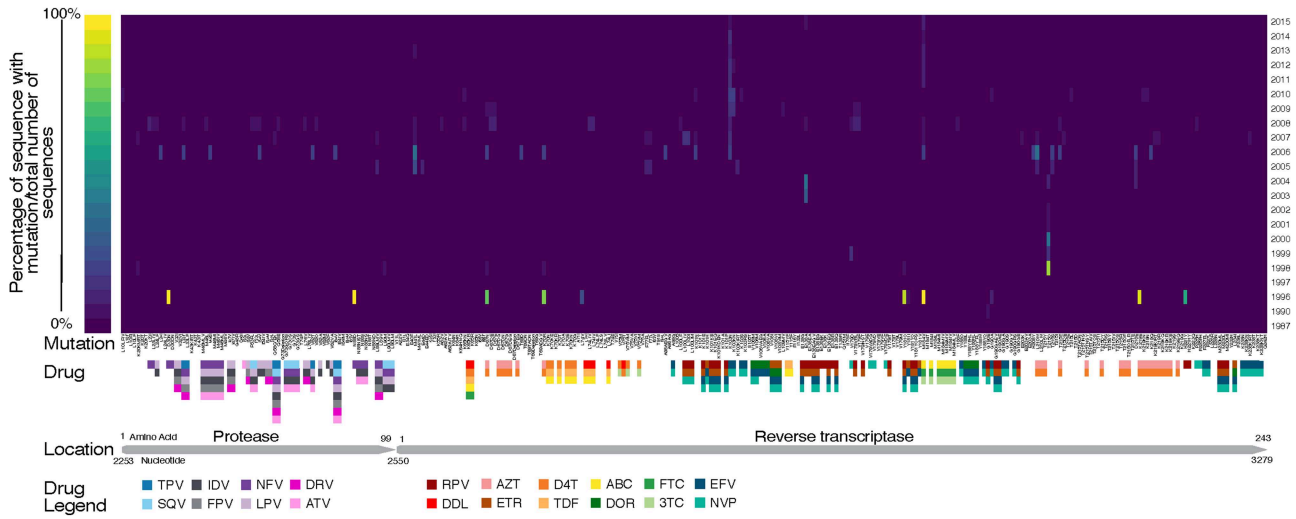


FIGURE 6 | Heat map of drug resistant mutations over time. Annotated with genome location, relative to HXB2 reference, and known drugs that the mutations are selected by. TPV, Tipranavir; SQV, Saquinavir; NFV, Nelfinavir; LPV, Lopinavir; IDV, Indinavir; DRV, Darunavir; ATV, Atazanavir; RPV, Rilpivirine; DDL, Didanosine; AZT, Zidovudine; ETR, Etravirine; D4T, Stavudine; TDF, Tenofovir Disoproxil Fumarate; ABC, Abacavir; DOR, Doravirine; FTC, Emtricitabine; 3TC, Lamivudine; EFV, Efavirenz; NVP, Nevirapine.

and MSM risk groups are primarily contributing to transmission of HIV (District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), 2017, 2018). As in our study, the majority of identified transmission clusters include at least one MSM and one HRH individual and often involve sequences obtained from the same year. Additionally, insights from DRM prevalence and genetic diversity trends over time will allow treatment measures to be more effective. Specifically, we found that both genetic diversity and the number of unique DRMs in circulation increased in subtype B sequences from more recent years. As such, our findings suggest that new surveillance studies of HIV subtypes should be conducted to better understand the current transmission dynamics of HIV in Washington, DC, especially in treatment-naïve individuals where our study and others (Kassaye et al., 2016; Ross et al., 2018) saw a spike between 2005 and 2008 in DRM.

DATA AVAILABILITY

The new HIV sequence data associated with this study have been deposited in GenBank under accession numbers MK270625 - MK270927.

ETHICS STATEMENT

Individuals represented in the new HIV data participated in clinical protocols (00-I-0110, 97-I-0082, and 95-I-0072), which took place at the NIH Clinical Center in Bethesda, Maryland. These protocols were approved by the NIAID Institutional Review Board (FWA00005897). All participants provided written informed consent.

AUTHOR CONTRIBUTIONS

KC, MP-L, ZG, FM, and SK designed the project. SK, ZG, and FM collected data. KG and MS analyzed the data and performed all bioinformatic analyses. MP-L and KC reviewed all analyses. KG, MS, MP-L, and KC prepared original draft of manuscript, and all authors read, revised and approved the final manuscript.

FUNDING

This study was partially supported by a Metropolitan Washington WIHS (Seble Kassaye, PI) U10-A1-034994 award, DC D-CFAR pilot award, USA NIH CCNS214443F and NIH UL1TR000075 awards from the NIH National Center for Advancing Translational Sciences, the DC Cohort Study (UO1 AI69503- 03S2), a supplement from the Women's Interagency Study for HIV (410722_GR410708), and a 2015 HIV Phylodynamics Supplement award from the District of Columbia Center for AIDS Research, an NIH funded program (P30AI117970), which is supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, NIGMS, NIDDK, and OAR. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

ACKNOWLEDGMENTS

We thank all the authors from the two manuscripts whose data was incorporated into this project (Maldarelli et al., 2013; Pérez-Losada et al., 2017) for their work on those respective projects, as their efforts were vital to the inclusion of those sequencing data.

REFERENCES

Arts, E. J., Hazuda, D. J., Bushman, E. F. D., Nabel, G. J., and Swanstrom, R. (2012). HIV-1 antiretroviral drug therapy basic principles of antiretroviral. *Cold Spring Harb. Perspect. Med.* 2, 1–23. doi: 10.1101/cshperspect.a007161

Bacheler, L. T., Anton, E. D., Kudish, P., Baker, D., Bunville, J., Krakowski, K., et al. (2000). Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. *Antimicrob. Agents Chemother.* 44, 2475–2484. doi: 10.1128/AAC.44.9.2475-2484.2000

Barré-Sinoussi, F., Ross, A. L., and Delfraissy, J. F. (2013). Past, present and future: 30 years of HIV research. *Nat. Rev. Microbiol.* 11, 877–883. doi: 10.1038/nrmicro3132

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, 1–6. doi: 10.1371/journal.pcbi.1003537

Centers for Disease Control and Prevention (2017). *HIV in the United States by Geography*. Available online at: https://www.cdc.gov/hiv/pdf/statistics/cdc-hiv-geographic-distribution.pdf.

Crandall, K. A., Posada, D., and Vasco, D. (1999). Effective population sizes: missing measures and missing concepts. *Anim. Conserv.* 2, 317–319. doi: 10.1111/j.1469-1795.1999.tb00078.x

Dalai, S. C., Junqueira, D. M., Wilkinson, E., Mehra, R., Kosakovsky Pond, S. L., Levy, V., et al. (2018). Combining phylogenetic and network approaches to identify hiv-1 transmission links in san mateo county, California. *Front. Microbiol.* 9, 1–8. doi: 10.3389/fmicb.2018.02799

DiNenno, E. A., Oster, A. M., Sionean, C., Denning, P., and Lansky, A. (2012). Piloting a system for behavioral surveillance among heterosexuals at increased risk of HIV in the United States. *Open AIDS J.* 6, 169–176. doi: 10.2174/1874613601206010169

District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), (2017). *Annual Epidemiology and Surveillance Report: Data Through December 2016*. Washignton, DC Available online at: www.doh.dc.gov/hahsta

District of Columbia Department of Health, HIV/AIDS, Hepatitis, STD and TB Administration (HAHSTA), (2018). *Annual Epidemiology and Surveillance Report: Data Through December 2017 District*. Available online at: www.doh.dc.gov/hahsta

Esbjörnsson, J., Mild, M., Audelin, A., Fonager, J., Skar, H., Bruun Jørgensen, L., et al. (2016). HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. *Virus Evol.* 2:vew010. doi: 10.1093/ve/vew010

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y)* 39, 783–791.

Frost, S. D., Nijhuis, M., Schuurman, R., Boucher, C. A. B., and Brown, A. J. L. (2000). Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *J. Virol.* 74, 6262–6268. doi: 10.1128/JVI.74.14.6262-6268.2000

Geneva: World Health Organization (2016). *Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection. Recommendations for a Public Health Approach. Second Edition.* Available online at: https://www.who.int/hiv/pub/arv/arv-2016/en/.

Geneva: World Health Organization (2017). *HIV Drug Resistance Report 2017*.

Germer, J. J., Wu, P., Soderberg, J. D., Mandrekar, J. N., and Yao, J. D. (2015). HIV-1 subtype diversity among clinical specimens submitted for routine antiviral drug resistance testing in the United States. *Diagn. Microbiol. Infect. Dis.* 83, 257–260. doi: 10.1016/j.diagmicrobio.2015.07.014

Greenberg, A. E., Hader, S. L., Masur, H., Young, A. T., Skillicorn, J., and Dieffenbach, C. W. (2009). Fighting HIV/AIDS in Washington, D.C. *Health Aff.* 28, 1677–1687. doi: 10.1377/hlthaff.28.6.1677

Grossman, Z., Rico, S. V., Cone, K., Shao, W., Rehm, C., Jones, S., et al. (2018). Early presence of HIV-1 subtype C in Washington, D.C. *AIDS Res. Hum. Retroviruses* 34, 680–684. doi: 10.1089/aid.2018.0041

Gulick, R. M., Ribaudo, H. J., Shikuma, C. M., Lustgarten, S., Squires, K. E., Meyer, W. A., et al. (2004). Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. *N. Engl. J. Med.* 350, 1850–1861. doi: 10.1056/NEJMoa031772

Ivanov, I. A., Beshkov, D., Shankar, A., Hanson, D. L., Paraskevis, D., Georgieva, V., et al. (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

Kassaye, S. G., Grossman, Z., Balamane, M., Johnston-White, B., Liu, C., Kumar, P., et al. (2016). Transmitted HIV drug resistance is high and longstanding in metropolitan Washington, DC. *Clin. Infect. Dis.* 63, 836–843. doi: 10.1093/cid/ciw382

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Keulen, W., Boucher, C., and Berkhout, B. (1996). Nucleotide substitution patterns can predict the requirements for drug-resistance of HIV-1 proteins. *Antivir. Res.* 31, 45–57. doi: 10.1016/0166-3542(96)00944-8

Kosakovsky Pond, S. L., Weaver, S., Leigh Brown, A. J., and Wertheim, J. O. (2018). HIV-TRACE (Transmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. *Mol. Biol. Evol.* 35, 1–16. doi: 10.1093/molbev/msy016

Latkin, C., Yang, C., Tobin, K., Penniman, T., Patterson, J., and Spikes, P. (2011). Differences in the social networks of African American men who have sex with men only and those who have sex with men and women. *Am. J. Public Health* 101, 18–23. doi: 10.2105/AJPH.2011.300281

Liu, T. F., and Shafer, R. W. (2006). Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin. Infect. Dis.* 42, 1608–1618. doi: 10.1086/503914

Magnus, M., Kuo, I., Shelley, K., Rawls, A., Peterson, J., Montanez, L., et al. (2009). Risk factors driving the emergence of a generalized heterosexual HIV epidemic in Washington, district of Columbia networks at risk. *Aids* 23, 1277–1284. doi: 10.1097/QAD.0b013e32832b51da

Maldarelli, F., Kearney, M., Palmer, S., Stephens, R., Mican, J., Polis, M. A., et al. (2013). HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *J. Virol.* 87, 10313–10323. doi: 10.1128/JVI.01225-12

Margot, N. A., Lu, B., Cheng, A., and Miller, M. D. (2006). Resistance development over 144 weeks in treatment-naive patients receiving tenofovir disoproxil fumarate or stavudine with lamivudine and efavirenz in Study 903. *HIV Med.* 7, 442–450. doi: 10.1111/j.1468-1293.2006.00404.x

Milken Institute School of Public Health (2018). *D.C. Cohort Longitudinal HIV Study*. Available online at: http://go.gwu.edu/dccohort.

Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyride through a rough skyline: bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25, 1459–1471. doi: 10.1093/molbev/msn090

Morgan, E., Oster, A. M., Townsell, S., Peace, D., Benbow, N., and Schneider, J. A. (2017). HIV-1 infection and transmission networks of younger people in Chicago, Illinois, 2005-2011. *Public Health Rep.* 132, 48–55. doi: 10.1177/0033354916679988

Neogi, U., Häggblom, A., Santacatterina, M., Bratt, G., Gisslén, M., Albert, J., et al. (2014). Temporal trends in the Swedish HIV-1 epidemic: increase in non-B subtypes and recombinant forms over three decades. *PLoS ONE* 9:e99390. doi: 10.1371/journal.pone.0099390

Pérez-Losada, M., Castel, A. D., Lewis, B., Kharfen, M., Cartwright, C. P., Huang, B., et al. (2017). Characterization of HIV diversity, phylodynamics and drug resistance in Washington, DC. *PLoS ONE* 12: 0185644. doi: 10.1371/journal.pone.0185644

Pérez-Losada, M., Jobes, D. V., Sinangil, F., Crandall, K. A., Posada, D., and Berman, P. W. (2010). Phylodynamics of HIV-1 from a phase-III AIDS vaccine trial in North America. *Mol. Biol. Evol.* 27, 417–425. doi: 10.1093/molbev/msp254

Pineda-Peña, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., et al. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* 19, 337–348. doi: 10.1016/j.meegid.2013.04.032

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Posada, D., and Crandall, K. (2001). Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol. Biol. Evol.* 18, 897–906. doi: 10.1093/oxfordjournals.molbev.a003890

Rambaut, A., Suchard, M., Xie, D., and Drummond, A. (2018). *Tracer* v1.6. Available online at: http://beast.bio.ed.ac.uk/%0A24.

Raymond, H. F., Al-Tayyib, A., Neaigus, A., Reilly, K. H., Braunstein, S., Brady, K. A., et al. (2017). HIV among MSM and heterosexual women in the United States: an ecologic analysis. *J. Acquir. Immune Defic. Syndr.* 75, S276–S280. doi: 10.1097/QAI.0000000000001422

Ross, L. L., Shortino, D., and Shaefer, M. S. (2018). Changes from 2000 to 2009 in the prevalence of HIV-1 containing drug resistance-associated mutations from antiretroviral therapy-naive, HIV-1-infected patients in the United States. *AIDS Res. Hum. Retroviruses* 34, 672–679. doi: 10.1089/aid.2017.0295

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu,033

U.S. Food and Drug Administration (2018). *HIV/AIDS Historical Time Line 2010–2017*. Silver Spring, MD.

UK Collaborative Group on HIV Drug Resistance (2014). The increasing genetic diversity of HIV-1 in the UK, 2002-2010. *AIDS* 28, 773–780. doi: 10.1097/QAD.0000000000000119

Vermund, S. H., and Leigh-Brown, A. J. (2012). The HIV epidemic: high-income countries. *Cold Spring Harb. Perspect. Med.* 2, 1–24. doi: 10.1101/cshperspect.a007195

Wertheim, J. O., Kosakovsky Pond, S. L., Forgione, L. A., Mehta, S. R., Murrell, B., Shah, S., et al. (2017). Social and genetic networks of HIV-1 transmission in New York City. *PLoS Pathog.* 13, 1–19. doi: 10.1371/journal.ppat.1006000

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 March 2019
doi: 10.3389/fmicb.2019.00370

Spatiotemporal Characteristics of the Largest HIV-1 CRF02_AG Outbreak in Spain: Evidence for Onward Transmissions

Evangelia-Georgia Kostaki[1], Andreas Flampouris[1], Timokratis Karamitros[2,3], Natalia Chueca[4], Marta Alvarez[4], Paz Casas[4], Belen Alejos[5], Angelos Hatzakis[1], Federico Garcia[4], Dimitrios Paraskevis[1] and CoRIS*

[1] Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece, [2] Department of Zoology, University of Oxford, Oxford, United Kingdom, [3] Public Health Laboratories, Department of Microbiology, Hellenic Pasteur Institute, Athens, Greece, [4] Department of Clinical Microbiology, Hospital Universitario San Cecilio, Instituto de Investigación Ibs, Granada, Spain, [5] Centro Nacional de Epidemiología, Universidad de Alcalá de Henares, Madrid, Spain

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Fernando Gonzalez-Candelas,
University of Valencia, Spain
Frank van der Meer,
University of Calgary, Canada

****Correspondence:***
Federico Garcia
fegarcia@ugr.es

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 October 2018*
Accepted: *12 February 2019*
Published: *11 March 2019*

Citation:
Kostaki E-G, Flampouris A, Karamitros T, Chueca N, Alvarez M, Casas P, Alejos B, Hatzakis A, Garcia F, Paraskevis D and CoRIS (2019) Spatiotemporal Characteristics of the Largest HIV-1 CRF02_AG Outbreak in Spain: Evidence for Onward Transmissions. Front. Microbiol. 10:370. doi: 10.3389/fmicb.2019.00370

Background and Aim: The circulating recombinant form 02_AG (CRF02_AG) is the predominant clade among the human immunodeficiency virus type-1 (HIV-1) non-Bs with a prevalence of 5.97% (95% Confidence Interval-CI: 5.41–6.57%) across Spain. Our aim was to estimate the levels of regional clustering for CRF02_AG and the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain.

Methods: We studied 396 CRF02_AG sequences obtained from HIV-1 diagnosed patients during 2000–2014 from 10 autonomous communities of Spain. Phylogenetic analysis was performed on the 391 CRF02_AG sequences along with all globally sampled CRF02_AG sequences (N = 3,302) as references. Phylodynamic and phylogeographic analysis was performed to the largest CRF02_AG monophyletic cluster by a Bayesian method in BEAST v1.8.0 and by reconstructing ancestral states using the criterion of parsimony in Mesquite v3.4, respectively.

Results: The HIV-1 CRF02_AG prevalence differed across Spanish autonomous communities we sampled from ($p < 0.001$). Phylogenetic analysis revealed that 52.7% of the CRF02_AG sequences formed 56 monophyletic clusters, with a range of 2–79 sequences. The CRF02_AG regional dispersal differed across Spain ($p = 0.003$), as suggested by monophyletic clustering. For the largest monophyletic cluster (subepidemic) (N = 79), 49.4% of the clustered sequences originated from Madrid, while most sequences (51.9%) had been obtained from men having sex with men (MSM). Molecular clock analysis suggested that the origin (t_{MRCA}) of the CRF02_AG subepidemic was in 2002 (median estimate; 95% Highest Posterior Density-HPD interval: 1999–2004). Additionally, we found significant clustering within the CRF02_AG subepidemic according to the ethnic origin.

Conclusion: CRF02_AG has been introduced as a result of multiple introductions in Spain, following regional dispersal in several cases. We showed that CRF02_AG transmissions were mostly due to regional dispersal in Spain. The hot-spot for the

largest CRF02_AG regional subepidemic in Spain was in Madrid associated with MSM transmission risk group. The existence of subepidemics suggest that several spillovers occurred from Madrid to other areas. CRF02_AG sequences from Hispanics were clustered in a separate subclade suggesting no linkage between the local and Hispanic subepidemics.

Keywords: HIV-1, CRF02_AG, Spain, regional dispersal, spatiotemporal characteristics

INTRODUCTION

Human immunodeficiency virus (HIV)-pandemic has been caused by Group M which is divided into nine subtypes (A, B, C, D, F, G, H, J, and K), six sub-subtypes (A1, A2, A3, A4, F1, F2) but also an extensive list of at least 98 CRFs which result from the recombination of two or more different subtypes (Foley et al., 2016)[1]. The distribution of subtypes and CRFs differ greatly across the globe. The majority (46.6%) of the infections worldwide are caused by subtype C, followed by subtype B (12.1%) and subtype A (10.3%); the former is predominant in the Western world (Hemelaar et al., 2018). CRF02_AG is the fourth worldwide (7.7%), however, it accounts for approximately 50% of the HIV-infections in West and Central Africa (Delatorre et al., 2014). CRF01_AE (5.3%), subtype G (4.6%) and D (2.7%) are following (Hemelaar et al., 2018).

The putative origin of CRF02_AG was in Central Africa, thereafter the virus spread into Western Africa establishing a major epidemic (Mir et al., 2016). Further dissemination from Western Africa occurred to Cameron in the late seventies, to the former Soviet Union in the late nineties and later to Bulgaria and Germany (Mir et al., 2016). CRF02_AG remains one of the most prevalent CRFs in Europe among individuals from highly endemic countries but also in non-migrant populations (Abecasis et al., 2013; Beloukas et al., 2016).

A previous study on 6,633 sequences sampled across Spain, revealed that subtype B was the most prevalent, but the non-B clades were dominated by CRFs. Specifically, it was found that CRF02_AG had a prevalence of 5.97% (95% CI: 5.41–6.57%) and was the most prevalent among the non-B clades (Flampouris et al., 2014). It is estimated that 130,000–160,000 people are living with HIV in Spain (Morán Arribas et al., 2018) while there is a high burden of late diagnosis. Current HIV epidemic in Spain is dominated by infections among MSM (del Amo and Iniesta, 2018).

The aim of the current study was to estimate the levels of regional clustering for CRF02_AG in Spain, to assess parameters associated with its regional dispersal across Spain, but also to estimate the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain. We also investigated whether regional dispersal is due to migration or due to onward transmissions among non-migrants.

Abbreviations: CI, confidence interval; CRFs, circulating recombinant forms; CRF02_AG, circulating recombinant form 02_AG; HIV-1, human immunodeficiency virus type-1; MCMC, Markov chain Monte Carlo; MSM, men having sex with men; PLHIV, people living with HIV; t_{MRCA}, time to the most recent common ancestor.

[1]https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html

MATERIALS AND METHODS

Study Population

We studied 396 HIV-1 CRF02_AG sequences available in the protease and partial reverse transcriptase regions of the virus genome (*pol gene*). Sequences were obtained from HIV-1 diagnosed patients during 2000–2014 from 10 autonomous communities of Spain. Study population characteristics are shown in **Table 1**. Specifically, patients' samples were merged from two datasets: (a) a representative cohort of HIV-infected patients included in the Research Network on HIV/AIDS (CoRIS) (2004–2013) and (b) Eastern Andalusia Resistance Cohort (2000–2014). A detailed description of the CoRIS cohort has been published previously (Caro-Murillo et al., 2007; Yebra et al., 2012). For this study, fully anonymized nucleotide sequences were retrospectively analyzed thus no written informed consent is required. This research project has been approved by the HU san Cecilio's Ethics Committee.

DNA Sequence Alignment and Phylogenetic Analysis

The patterns of CRF02_AG dispersal in Spain were investigated by means of phylogenetic analysis. Analysis was performed in 391 out of 396 (98.7%) CRF02_AG sequences along with all globally sampled CRF02_AG sequences available on the public HIV-1 sequence database[2] (N = 3,302) as references. A small number of identical sequences (5 of 396, 1.3%) was excluded from analysis to avoid inclusion of duplicates. MEGA v5.2 (Hall, 2013) was used to align sequences using the MUSCLE algorithm and alignments were manually edited according to the encoded reading frame (only codons and no single or double combinations of nucleotides were excluded from the alignment). Codons associated with resistance (IAS mutation list 2017) were excluded to avoid potentially bias on clustering due to convergent evolution at resistance sites (Lewis et al., 2008). The final alignment was consisted of 738 nucleotides. FigTree v1.4 was used for tree visualization and annotation[3].

The phylogenetic trees were estimated from the underlying nucleotide sequences. Initially, Maximum likelihood phylogeny reconstruction with bootstrap evaluation was conducted in RAxML v8.0.20 (Stamatakis, 2014) using the general time-reversible (GTR) substitution model and gamma (Γ) distribution. Subsequently, further analysis was performed on the clusters that initially received bootstrap value lower than 75%, by using

[2]https://www.hiv.lanl.gov

[3]http://tree.bio.ed.ac.uk/software/figtree/

a well-justified method based on Bayesian analysis (Kouyos et al., 2010; Paraskevis et al., 2019). In detail, we used the Simple Consensus Maker algorithm[4] to make the consensus sequence of each cluster with bootstrap value lower than the threshold. The 100 most closely related sequences were found by ordering the consensus sequence with all the CRF02_AG available sequences (391 sequences from our study population and 3,302 reference sequences) according to their similarity, in Mafft v7.4 (Katoh et al., 2017). Sequences found within each cluster were then analyzed phylogenetically along with the 100 most closely related sequences to them, using the Bayesian method with the GTR substitution model with Γ distributed rate, as implemented in MrBayes v3.2.2 (Ronquist and Huelsenbeck, 2003). The MCMC ran for 10×10^5 generations (burn-in: 10%), with 4 chains per run, and with MCMC sampling every 1,000 steps. Each MCMC run checked for convergence using Tracer version v1.5 (Drummond et al., 2012). Thereafter, phylogenetic clusters were defined as monophyletic using two different criteria: (i) clusters with bootstrap values greater than 75%, for phylogenetic trees estimated by the Maximum likelihood method, or a posterior probability greater than 0.82[5], for phylogenetic trees estimated by the Bayesian method (phylogenetic confidence criterion), and (ii) clusters consisting of at least two sequences sampled from Spain at a proportion greater than 70% compared to total number of sequences within the cluster (geographic criterion). Only phylogenetic clusters fulfilling both criteria were considered as monophyletic.

[4]https://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html

TABLE 1 | Characteristics of the study population: (i) total population (*N* = 396), and (ii) a subset of the total population (*N* = 391 of 396) after the exclusion of identical sequences.

Characteristic	Total population	Subset of the total population
Median age [years (IQR[1])]	35.6 (29.3–43.6)[2]	36.2 (29.4–43.6)[3]
Gender (N, %)		
Male	233 (58.8)	230 (58.8)
Female	163 (41.2)	161 (41.2)
Transmission risk group (N, %)		
Heterosexuals	158 (39.9)	156 (39.9)
MSM[4]	62 (15.7)	60 (15.4)
PWID[5]	1 (0.2)	1 (0.2)
Other/unknown	175 (44.2)	174 (44.5)
Spanish autonomous community of sampling (N, %)		
Andalusia	240 (60.6)	239 (61.3)
Madrid	76 (19.2)	75 (19.2)
Valencia	22 (5.6)	19 (4.9)
Catalonia	18 (4.6)	18 (4.6)
La Rioja	11 (2.8)	11 (2.8)
Navarre	11 (2.8)	11 (2.8)
Basque Country	11 (2.8)	11 (2.8)
Balearic Islands	3 (0.7)	3 (0.7)
Galicia	3 (0.7)	3 (0.7)
Canary Islands	1 (0.2)	1 (0.2)
Total (N, %)	**396 (100)**	**391 (100)**

[1]IQR, interquartile range; [2]N = 313 (of 396 individuals); [3]N = 309 (of 391 individuals); [4]MSM, men having sex with men; [5]PWID, people who inject drugs.

Phylodynamic Analysis

We further conducted phylodynamic analysis to estimate the spatiotemporal characteristics of the largest CRF02_AG subepidemic (monophyletic cluster consisted of 79 sequences). Analysis was performed only on sequences with available information about their date of sampling (72 of 79, 91.1%). We performed the analysis by a Bayesian method using the HKY nucleotide substitution model, a Γ distributed rate of heterogeneity among sites, an uncorrelated log normal relaxed clock of molecular clock model with TipDates and the birth-death basic reproductive number (R_e) models as implemented in BEAST v1.8.0 (Drummond et al., 2012). Non-informative priors were used for the MCMC runs. MCMC analysis was run for 30×10^6 generations, sampled every 3.000 steps (burn-in: 10%). The MCMC convergence was checked using Tracer v1.5 (Drummond et al., 2012), by estimating the effective sample sizes (ESS > 200). The temporal signal in our data was tested by TempEst (Rambaut et al., 2016).

Phylogeographic Analysis

Phylogeographic analysis was inferred by character reconstruction using the criterion of parsimony on the dated phylogeny using Mesquite v3.4[6].

Statistical Analysis

Demographic data are summarized using median and interquartile ranges for continuous variables, and absolute and relative frequencies for categorical variables. Statistical analysis for simple comparisons of the relevant distributions across different levels of other categorical variables was carried out using Pearson's chi-squared statistical test. A multivariate logistic regression model was fit to a subset of the original data (N = 391), consisting of 304 complete observations. Presence in monophyletic groups was the binary outcome variable, while age, gender, transmission risk group and Spanish autonomous community of sampling were chosen as possible explanatory variables. Statistical analysis was performed on STATA 12-StataCorp LP.

RESULTS

CRF02_AG prevalence was significantly different across the Spanish autonomous communities we sampled from ($p < 0.001$).

[5]All clusters receive posterior probability support >0.82, expect for two clusters consisting of two sequences each, for which the posterior probability support was 0.76 and 0.79. The latter were included in the analysis since the confirmatory analysis was based on stringent sampling criteria.

[6]http://www.mesquiteproject.org

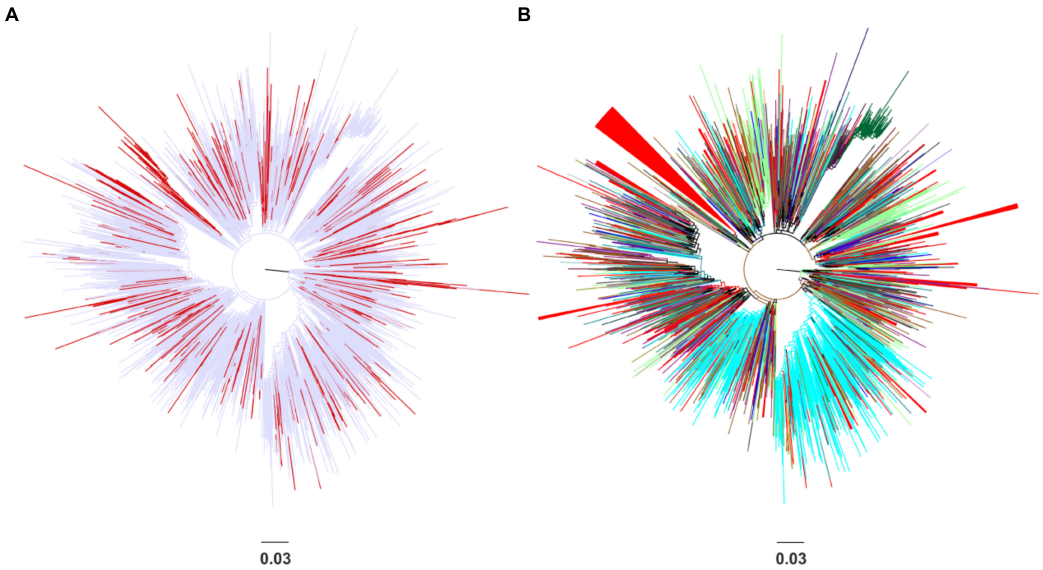


FIGURE 1 | Unrooted phylogenetic trees estimated by RAxML v8.0.20 of HIV-1 CRF02_AG sequences sampled from Spain and a global reference dataset. **(A)** Sequences from Spain are marked in red in contrast with sequences from other areas marked in light blue. **(B)** Sequences from Spain are marked in red in contrast with sequences from other countries (Cameroon, Senegal, Niger, Ghana, Mali, the United States, Burkina Faso, Togo, Italy, Russia, Benin, Gabon, Germany, Belgium, the United Kingdom, Japan, France, Mauritania, Portugal, Sweden, Equatorial Guinea, the Czech Republic, Denmark, Cyprus, Australia, and Canada) and geographical areas (Western Europe, Sub-Saharan Africa, Central Asia, East Asia, Central and Eastern Europe, Southeast Asia, and Latin America) marked in different colors. Monophyletic clusters according to geographic region of sampling are indicated as triangles.

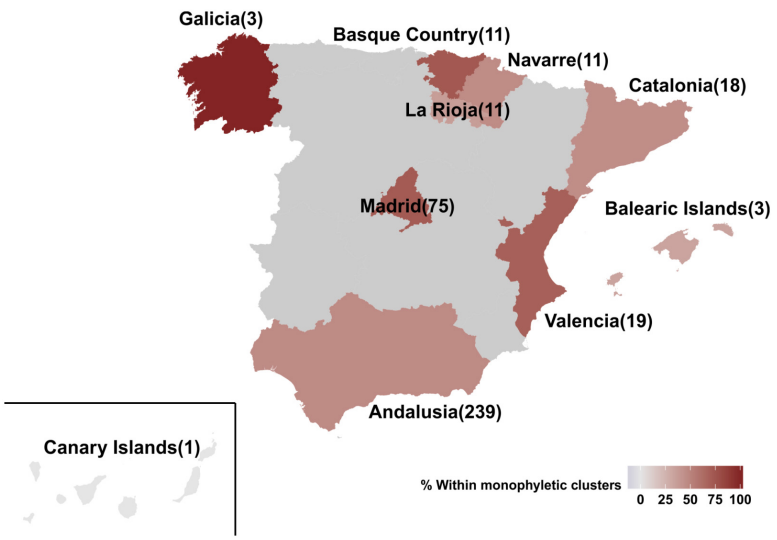


FIGURE 2 | Map of Spain representing the 10 autonomous communities from which the 391 HIV-1 CRF02_AG sequences of our study population were sampled from. Autonomous communities are colored according to the percentage of monophyletic clustering. Numbers in brackets describe the available sequences per autonomous community.

La Rioja (N = 11, 16.7%, 95% CI: 8.6–27.9%), Navarre (N = 11, 13.6%, 95% CI: 7–23%), Valencia (N = 22, 7.3%, 95% CI: 5.3–12.4%), and Andalusia (N = 240, 7%, 95% CI: 6.2–8.0%) were the autonomous communities where CRF02_AG was most prevalent.

Phylogenetic analysis revealed that the CRF02_AG sequences sampled from Spain clustered at different points in the Maximum-likelihood tree (**Figure 1A**). No specific geographic area was identified as the origin of CRF02_AG in Spain, since the Spanish monophyletic clusters appeared as nested within sequences from different countries (**Figure 1B**). The high-levels of dispersal of the CRF02_AG strains (red branches) across the tree (**Figure 1A**) or the existence of a large number of monophyletic clusters (**Figure 1B**) indicate that the introduction

of CRF02_AG into Spain has been occurred from multiple sources. Specifically, we found that 52.7% (206 of 391) of our study CRF02_AG sequences formed 56 monophyletic clusters, with a range of 2–79 sequences (**Figure 1B**). The median number of sequences per monophyletic cluster was 3 (Interquartile range, IQR: 2–5). Phylogenetic analysis revealed the presence of highly supported phylogenetic clusters consisted of sequences from Spain at proportions >70% (monophyletic clusters). HIV-1 sequences found within the monophyletic clusters suggest a common route of infection. The fact that the majority of sequences have been retrieved from PLHIV residing in Spain implies that monophyletic clusters correspond to transmissions occurred locally (regional transmissions).

The percentage of regional dispersal differed across Spain, as suggested by the monophyletic clustering ($p = 0.003$) (**Figure 2**). Specifically, the highest proportion of regional dispersal was found in Basque Country where 72.7% (8 of 11) of sequences belonged to monophyletic clusters, followed by sequences from Madrid (54 of 75, 72%) and Valencia (13 of 19, 68.4%). Sequences from Andalusia (110 of 239, 46%), Navarre (5 of 11, 45.5%), Catalonia (8 of 18, 44.4%), and La Rioja (4 of 11, 36.4%) showed the lowest monophyly levels. For Galicia ($N = 3$), Canary Islands ($N = 1$), and Balearic Islands ($N = 3$) less than 10 sequences were available. The current study was performed using data from the representative cohort CoRIS (Caro-Murillo et al., 2007; Yebra et al., 2012; Vourli et al., 2019) and the Eastern Andalusia Resistance Cohort. The study population included sequences sampled during 2000–2014 from 10 autonomous communities of Spain, suggesting high geographic coverage. The proportion of sequences within each cluster was comparable across different communities, expect for Andalusia which was overrepresented. Despite its over-representativeness, Andalusia was found among the communities with the lowest levels of monophyly.

Using a multivariate logistic regression model, we found that the MSM transmission risk group was positively associated with the regional clustering ($p < 0.001$), having adjusted for the rest of the incorporated variables (**Table 2**). This association was mostly due to the large monophyletic cluster in which the majority of PLHIV were MSM.

The largest CRF02_AG monophyletic cluster (subepidemic) consisted of 79 sequences (**Figure 1B**). We found that 49.4% (39 of 79) of the clustered sequences originated from Madrid and most sequences (41 of 79, 51.9%) had been obtained from MSM. Phylogenetic analysis also revealed the existence of two nested clusters from Japan ($N = 7$, 8.9%) and Sweden ($N = 3$, 3.8%). The nested clusters consisted of reference sequences that were available in the HIV-1 sequence database.

To describe the temporal patterns of the largest CRF02_AG regional epidemic in Spain, we carried out phylodynamic analysis. There was evidence for temporal signal in the sequences found within the largest CRF02_AG monophyletic cluster, as tested by the TempEst program ($R = 0.721$). Moreover, the temporal structure in this data was assessed by estimating the ESS for all parameters of the MCMC analysis. In the MCMC analysis the ESS for all parameters were >>200. As explained in detail in the methods section, phylodynamic analysis was performed only on sequences with available information about their date of sampling ($N = 72$). For 57 out of 72 (79.2%) sequences found within the largest CRF02_AG monophyletic cluster, the sampling area was in Spain (**Table 3**). Molecular clock analysis suggested that the time to most recent common ancestor (t_{MRCA}) of the CRF02_AG subepidemic was in 2002 (median estimate; 95% Highest Posterior Density-HPD interval: 1999–2004). The t_{MRCA} should be considered as the approximate time of infection of the potential founder of the CRF02_AG subepidemic in Spain sampled in our data. The Bayesian skyline plot showed that the CRF02_AG subepidemic growth occurred during a period of 8 years and specifically between 2003 and 2011 (**Figure 3**).

TABLE 2 | Multivariate logistic regression estimates using the presence in monophyletic groups as the binary outcome variable.

Explanatory variable	Odds ratio	95% Conf. interval	*P*-value
Age	1.01	(0.99, 1.03)	0.506
Gender (*Male)			
Female	1.09	(0.65, 1.84)	0.732
Transmission risk group (*Heterosexuals)			
MSM[1]	6.66	(2.57, 17.23)	<0.001
PWID[2]	1	–	–
Other/unknown	0.63	(0.34–1.18)	0.153
Spanish autonomous community of sampling (*Andalusia)			
Madrid	0.88	(0.41, 1.91)	0.749
Valencia	1.04	(0.33, 3.22)	0.948
Catalonia	0.45	(0.15, 1.34)	0.152
La Rioja	0.48	(0.12, 1.90)	0.299
Navarre	0.47	(0.12, 1.82)	0.277
Basque Country	1.19	(0.26, 5.29)	0.823
Balearic Islands	0.33	(0.03, 3.93)	0.381
Galicia	1	–	–
Canary Islands	1	–	–

*The model was fit to a subset of the original data (N = 391), consisting of 304 complete observations. *Reference category: [1]MSM, men having sex with men; [2]PWID, people who inject drugs.*

TABLE 3 | Sampling region/country for sequences with known date of sampling found within the largest CRF02_AG monophyletic cluster (subepidemic) in Spain.

Region of sampling	Country of sampling	Number of sequences per country (%)	Number of sequences per region (%)
Western Europe	Spain	57 (79.2)	65 (90.3)
	Switzerland	4 (5.6)	
	Sweden	3 (4.2)	
	Germany	1 (1.2)	
East Asia	Japan	7 (9.8)	7 (9.7)
Total		**72 (100)**	**72 (100)**

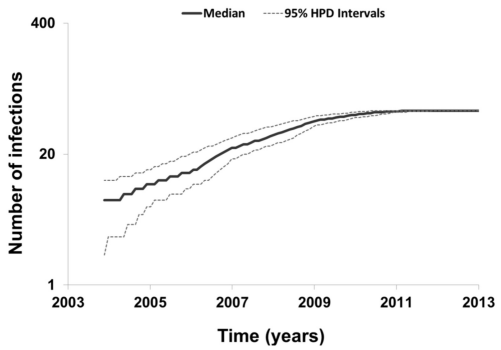


FIGURE 3 | Bayesian skyline plot estimated by BEAST v1.8.0 using birth-death models for the largest HIV-1 CRF02_AG subepidemic in Spain, showing the cumulative number of lineages (infections) in a logarithmic scale over time (median and 95% Highest Posterior Density-HPD intervals estimates).

Phylogeographic analysis revealed significant clustering within the CRF02_AG subepidemic according to the ethnic origin (**Figure 4A**). Specifically, four distinct subclades were found. The subclade division was based on high posterior probability support (>0.85) and epidemiologic criteria (i.e., the ethnic origin of PLHIV within each cluster). The first subclade was consisted of sequences from Spain (N = 44) including as nested the cluster from Japan (N = 7) (subclade I) (**Figures 4A,B**). The majority of Spanish lineages were from individuals living in Madrid (N = 30, 68.2%) reported MSM (N = 37, 84.1%) as transmission risk group (**Figures 4B,C**). Except from Madrid, Spanish sequences were obtained across five different autonomous communities of Spain (**Figure 4B**). The second subclade was consisted of sequences from Hispanics from Spain (N = 1, 17%) and Latin America (N = 5, 83%) (subclade II) (**Figure 4A**). The third subclade consisted of sequences (N = 6) originating from Western Europe (Sweden, Switzerland, and Germany) (subclade III) and the fourth of sequences (N = 4) from Colombia, Spain, and Switzerland (subclade IV) (**Figure 4A**). Subclades III and IV were more genetically divergent than the two others (as suggested from the within-subcluster branch lengths) (**Figure 4A**). Notably, the geographic origin of the CRF02_AG clade including sequences from Spain (subclades I and II) was estimated in Madrid, from where it spread to other autonomous communities within the country as well as outside Spain (**Figures 4A,B**).

DISCUSSION

In this study we used a collection of sequences sampled from 10 Spanish autonomous communities and a combination of phylogenetic, phylogeographic and phylodynamic analyses, in order to investigate the characteristics of the HIV-1 CRF02_AG epidemic in this country. We further shed light on the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain. Our analysis suggests that CRF02_AG has been introduced as a result of multiple introductions in Spain, following regional dispersal in several cases. However, we found considerable variation in the patterns of regional clustering in Spain with Basque area, Madrid and Valencia to be the areas with the higher proportion of regional transmissions. We estimated that the epidemic growth of the largest monophyletic cluster continued until 2011. The hub for the largest cluster was in Madrid from where cross-border and local transmissions to other autonomous communities were monitored. Notably, CRF02_AG sequences within the largest cluster in Madrid were detected as far as from Japan, comprising a local epidemic, as well as from Latin America. Although an alternative hypothesis cannot be

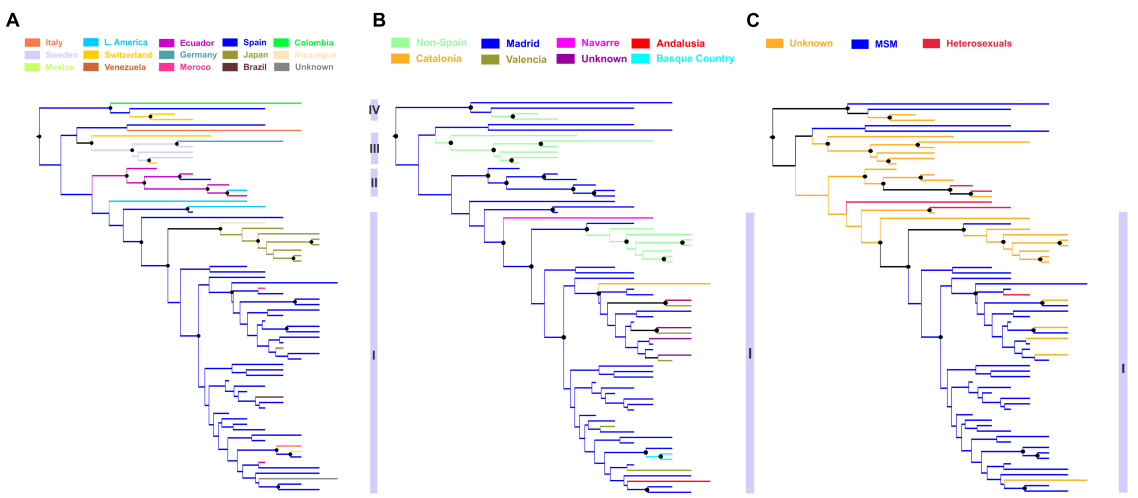


FIGURE 4 | Phylogeographic topology estimated by Mesquite v3.4 for largest HIV-1 CRF02_AG subepidemic in Spain, showing the viral clades in different colors according to the: **(A)** ethnic origin of the HIV-1 diagnosed patient, **(B)** Spanish autonomous community of sampling, and **(C)** transmission risk group of the HIV-1 diagnosed patient. Branch lengths were longer within subclusters III, IV than within subclusters I and II, suggesting higher genetic diversity. Black disks indicate nodes with posterior probability greater than 0.85.

entirely excluded, our suggestion that the Japanese local epidemic originated from Spain provides the most plausible hypothesis given the existing sampling. The long-distance dispersal of CRF02_AG is not strange since some of the sequences were obtained from Hispanic populations outside Spain, a finding that can be reconciled because of the historical links between Spain and Hispanic populations of the Americas. On the other hand, the high popularity of Spain as a travel destination could explain the outgoing transmissions to Japan. CRF02_AG regional dispersal in Spain was associated with MSM; a finding that was probably due to the pattern found for the largest cluster in Madrid.

The identification of Madrid as a hub for the largest cluster of CRF02_AG in Spain and that only small clusters were found outside Madrid, suggest that transmission networking is probably more extensive in the capital than in peripheral cities. Notably, similar findings were reported for Paris and Rome which provided as the source for the CRF02_AG transmissions across France and Italy, respectively (Giuliani et al., 2013; Chaillon et al., 2017). The identity of capital cities as hubs for within country dispersal could be explained by the fact that they provide social, economic cultural and recreational centers as well as major hubs for international traveling. These findings could be useful for public health stakeholders to focus their preventive actions in large cities in Europe.

The prevalence of CRF02_AG has been estimated to 3.0% in Western and Central Europe and North America, being the most frequently circulated among the CRFs (Hemelaar et al., 2018). Although sporadic CRF02_AG cases have been increasingly detected in Europe and North America as a result of population mobility from endemic regions, regional dispersal was also found in several areas in Europe including MSM and heterosexual populations in France, Belgium, Switzerland, and Spain (Dauwe et al., 2015), as well as people who inject drugs in Bulgaria (Alexiev et al., 2016). Specifically, only one small cluster was found in Belgium, while 28% of the CRF02_AG sequences belonged to Swiss-specific subclusters (Von Wyl et al., 2011). In France, 20.4% of the PLHIV primarily infected with CRF02_AG belonged to transmission networks based on analysis using a genetic distance cut-off (Chaillon et al., 2017). This proportion was the highest for all HIV-1 subtypes (Chaillon et al., 2017). A large monophyletic cluster of CRF02_AG was also monitored among PLHIV with primary infection in Southeastern France (Tamalet et al., 2015). Similarly, the proportion of CRF02_AG sequences within monophyletic clusters was reported: 29.8, 27.3, 25.4, and 18.2% for Germany, Norway, Austria, and Italy, respectively (Paraskevis et al., 2019). These findings suggest a common pattern of CRF02_AG dispersal in Europe, where although CRF02_AG have been associated with highly endemic areas (Abecasis et al., 2013) as their putative source, a considerable proportion of their sequences fall within monophyletic clusters. It remains unclear, however, whether onward transmissions in Europe occur among migrant or non-migrant populations.

In the current study using a state-of-the-art molecular epidemiology approach, we were able to describe the spatiotemporal characteristics of the CRF02_AG infection in Spain and to investigate the nature of onward transmissions across different ethnic groups. Notably, we found that the majority of CRF02_AG sequences within the largest cluster were retrieved from non-migrant populations, suggesting that the within the country infections are due to onward transmissions among non-migrants. To our knowledge this is one of the few studies providing evidence that local dispersal of non-B subtypes is associated with non-migrants. These findings can be important for public health and particularly for the design of targeted interventions for the populations at the higher risk.

CENTERS AND INVESTIGATORS INVOLVED IN CoRIS

Executive Committee

Santiago Moreno, Inma Jarrín, David Dalmau, Maria Luisa Navarro, Maria Isabel González, Jose Luis Blanco, Federico Garcia, Rafael Rubio, Jose Antonio Iribarren, Félix Gutiérrez, Francesc Vidal, Juan Berenguer, Juan González.

Fieldwork, data management and analysis: Belén Alejos, Victoria Hernando, Cristina Moreno, Carlos Iniesta, Luis Miguel Garcia Sousa, Nieves Sanz Perez.

BioBanK HIV

Hospital General Universitario Gregorio Marañón: M Ángeles Muñoz-Fernández, Isabel María García-Merino, Irene Consuegra Fernández, Coral Gómez Rico, Jorge Gallego de la Fuente, Paula Palau Concejo.

Participating Centres

Hospital General Universitario de Alicante (Alicante):

Joaquín Portilla, Esperanza Merino, Sergio Reus, Vicente Boix, Livia Giner, Carmen Gadea, Irene Portilla, María Pampliega, Marcos Díez, Juan Carlos Rodríguez, José Sánchez-Payá.

Hospital Universitario de Canarias (San Cristobal de la Laguna): Juan Luis Gómez, Jehovana Hernández, María Remedios Alemán, María del Mar Alonso, María Inmaculada Hernández, Felicitas Díaz-Flores, Dácil García, Ricardo Pelazas, Ana López Lirola.

Hospital Universitario Central de Asturias (Oviedo): José Sanz Moreno, Alberto Arranz Caso, Cristina Hernández Gutiérrez, María Novella Mena.

Hospital Universitario 12 de Octubre (Madrid): Rafael Rubio, Federico Pulido, Otilia Bisbal, Asunción Hernando, Lourdes Domínguez, David Rial Crestelo, Laura Bermejo, Mireia Santacreu.

Hospital Universitario de Donostia (Donostia-San Sebastián): José Antonio Iribarren, Julio Arrizabalaga, María José Aramburu, Xabier Camino, Francisco Rodríguez-Arrondo, Miguel Ángel von Wichmann, Lidia Pascual Tomé, Miguel Ángel Goenaga, Mª Jesús Bustinduy, Harkaitz Azkune, Maialen Ibarguren, Aitziber Lizardi, Xabier Kortajarena.

Hospital General Universitario De Elche (Elche): Félix Gutiérrez, Mar Masiá, Sergio Padilla, Andrés Navarro, Fernando

Montolio, Catalina Robledano, Joan Gregori Colomé, Araceli Adsuar, Rafael Pascual, Marta Fernández, Elena García, José Alberto García, Xavier Barber.

Hospital General Universitario Gregorio Marañón (Madrid): Juan Berenguer, Juan Carlos López Bernaldo de Quirós, Isabel Gutiérrez, Margarita Ramírez, Belén Padilla, Paloma Gijón, Teresa Aldamiz-Echevarría, Francisco Tejerina, Francisco José Parras, Pascual Balsalobre, Cristina Diez, Leire Pérez Latorre.

Hospital Universitari de Tarragona Joan XXIII (Tarragona): Francesc Vidal, Joaquín Peraire, Consuelo Viladés, Sergio Veloso, Montserrat Vargas, Miguel López-Dupla, Montserrat Olona, Anna Rull, Esther Rodríguez-Gallego, Verónica Alba.

Hospital Universitario y Politécnico de La Fe (Valencia): Marta Montero Alonso, José López Aldeguer, Marino Blanes Juliá, María Tasias Pitarch, Iván Castro Hernández, Eva Calabuig Muñoz, Sandra Cuéllar Tovar, Miguel Salavert Lletí, Juan Fernández Navarro, Jose Miguel Molina.

Hospital Universitario La Paz/IdiPAZ: Juan González-garcia, Francisco Arnalich, José Ramón Arribas, Jose Ignacio Bernardino de la Serna, Juan Miguel Castro, Luis Escosa, Pedro Herranz, Victor Hontañón, Silvia García-Bujalance, Milagros García López-Hortelano, Alicia González-Baeza, Maria Luz Martín-Carbonero, Mario Mayoral, Maria Jose Mellado, Rafael Esteban Micán, Rocio Montejano, María Luisa Montes, Victoria Moreno, Ignacio Pérez-Valero, Berta Rodés, Talia Sainz, Elena Sendagorta, Natalia Stella Alcáriz, Eulalia Valencia.

Hospital San Pedro Centro de Investigación Biomédica de La Rioja (CIBIR) (Logroño): José Ramón Blanco, José Antonio Oteo, Valvanera Ibarra, Luis Metola, Mercedes Sanz, Laura Pérez-Martínez.

Hospital Universitari MutuaTerrassa (Terrasa): David Dalmau, Angels Jaén, Montse Sanmartí, Mireia Cairó, Javier Martinez-Lacasa, Pablo Velli, Roser Font, Mariona Xercavins, Noemí Alonso.

Complejo Hospitalario de Navarra (Pamplona) María Rivero, Jesús Repáraz, María Gracia Ruiz de Alda, María Teresa de León Cano, Beatriz Pierola Ruiz de Galarreta.

Hospital Universitario de La Princesa (Madrid): Ignacio de los Santos, Jesús Sanz Sanz, Ana Salas Aparicio, Cristina Sarriá Cepeda, Lucio Garcia-Fraile Fraile, Enrique Martín Gayo.

Hospital Universitario Ramón y Cajal (Madrid): Santiago Moreno, José Luis Casado, Fernando Dronda, Ana Moreno, María Jesús Pérez Elías, Cristina Gómez Ayerbe, Carolina Gutiérrez, Nadia Madrid, Santos del Campo Terrón, Paloma Martí, Uxua Ansa, Sergio Serrano, María Jesús Vivancos.

Hospital General Universitario Reina Sofía (Murcia): Enrique Bernal, Alfredo Cano, Antonia Alcaraz García, Joaquín Bravo Urbieta, Ángeles Muñoz, Maria Jose Alcaraz, Maria del Carmen Villalba.

Hospital Universitario San Cecilio (Granada): Federico García, José Hernández, Alejandro Peña, Leopoldo Muñoz, Paz Casas, Marta Alvarez, Natalia Chueca, David Vinuesa, Clara Martinez-Montes, Fernando García, Carlos Guerrero-Beltran.

Centro Sanitario Sandoval (Madrid): Jorge Del Romero, Carmen Rodríguez, Teresa Puerta, Juan Carlos Carrió, Mar Vera, Juan Ballesteros, Oskar Ayerdi.

Hospital Clínico Universitario de Santiago (Santiago de Compostela): Antonio Antela, Elena Losada, Antonio Aguilera.

Hospital Universitario Son Espases (Palma de Mallorca): Melchor Riera, María Peñaranda, María Leyes, Mª Angels Ribas, Antoni A Campins, Carmen Vidal, Francisco Fanjul, Javier Murillas, Francisco Homar.

Hospital Universitario Virgen de la Victoria (Málaga): Jesús Santos, Crisitina Gómez Ayerbe, Isabel Viciana, Rosario Palacios, Carmen María González.

Hospital Universitario Virgen del Rocío (Sevilla): Pompeyo Viciana, Nuria Espinosa, Luis Fernando López-Cortés.

Hospital Universitario de Bellvitge (Hospitalet de Llobregat): Daniel Podzamczer, Elena Ferrer, Arkaitz Imaz, Juan Tiraboschi, Ana Silva, María Saumoy.

Hospital Costa del Sol (Marbella): Julián Olalla, Alfonso del Arco, Javier de la torre, José Luis Prada, José María García de Lomas Guerrero, Javier Pérez Stachowski.

Hospital Marina Baixa (VilaJollosa, Alicante): Concepción Amador.

Hospital General Universitario Santa Lucía (Cartagena): Onofre Juan Martínez, Francisco Jesús Vera, Lorena Martínez, Josefina García, Begoña Alcaraz, Amaya Jimeno.

Complejo Hospitalario Universitario a Coruña (Chuac) (A Coruña): Angeles Castro Iglesias, Berta Pernas Souto, Alvaro Mena de Cea.

Hospital Universitario Virgen de la Arrixaca (El Palmar): Carlos Galera, Helena Albendin, Aurora Pérez, Asunción Iborra, Antonio Moreno, Maria Angustias Merlos, Asunción Vidal.

Hospital Universitario Infanta Sofia (San Sebastian de los Reyes): Inés Suárez-García, Eduardo Malmierca, Patricia González-Ruano, Dolores Martín Rodrigo.

Complejo Hospitalario de Jaén (Jaén): Mohamed Omar Mohamed-Balghata, María Amparo Gómez Vidal. Nuestra Señora de Valme: Juan A Pineda, Juan Macías, Samuel Bernal. Unidad de Biología y Variabilidad del VIH.

Centro Nacional de Microbiología, ISCIII: Miguel Thomson, Elena Delgado, Sonia Benito, Vanessa Montero.

DATA AVAILABILITY

The HIV-1 nucleotide sequences of our study population are available in GenBank with accession numbers MK588005–MK588395.

AUTHOR CONTRIBUTIONS

E-GK did the analysis and prepared the figures and the manuscript. AF and TK organized the data and built a database. NC, MA, PC, BA, and FG participated in the study design, performed the data collection across Spain, and provided critical comments about the manuscript. AH contributed to the study design and provided critical comments about the manuscript. DP performed the study design and supervised, and contributed to manuscript writing and editing.

FUNDING

This work was supported by Plan Nacional de I+D+I and Fondo Europeo de Desarrollo Regional-FEDER (RD16/0025/0040 and PI18/00819) and Fundación Progreso y Salud, Junta de Andalucia (PI-0550-2017). The study was in part funded by the Hellenic Scientific Society for the Study of AIDS and Sexually Transmitted Diseases.

REFERENCES

Abecasis, A. B., Wensing, A. M. J., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A., et al. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 10:7. doi: 10.1186/1742-4690-10-7

Alexiev, I., Shankar, A., Dimitrova, R., Gancheva, A., Kostadinova, A., Teoharov, P., et al. (2016). Origin and spread of HIV-1 in persons who inject drugs in Bulgaria. *Infect. Genet. Evol.* 46, 269–278. doi: 10.1016/j.meegid.2016.05.029

Beloukas, A., Psarris, A., Giannelou, P., Kostaki, E., Hatzakis, A., and Paraskevis, D. (2016). Molecular epidemiology of HIV-1 infection in Europe: an overview. *Infect. Genet. Evol.* 46, 180–189. doi: 10.1016/j.meegid.2016.06.033

Caro-Murillo, A. M., Castilla, J., Perez-Hoyos, S., Miro, J. M., Podzamczer, D., Rubio, R., et al. (2007). Spanish cohort of naïve HIV-infected patients (CoRIS): rationale, organization and initial results | Cohorte RIS de pacientes con infección por VIH sin tratamiento antirretroviral previo (CoRIS): Metodología y primeros resultados. *Enferm. Infecc. Microbiol. Clin.* 25, 23–31. doi: 10.1157/13096749

Chaillon, A., Essat, A., Frange, P., Smith, D. M., Delaugerre, C., Barin, F., et al. (2017). Spatiotemporal dynamics of HIV-1 transmission in France (1999-2014) and impact of targeted prevention strategies. *Retrovirology* 14, 1–12. doi: 10.1186/s12977-017-0339-4

Dauwe, K., Mortier, V., Schauvliege, M., Van Den Heuvel, A., Fransen, K., Servais, J. Y., et al. (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.* 15:524. doi: 10.1186/s12879-015-1217-0

del Amo, J., and Iniesta, C. (2018). Missed opportunities within the health system for the diagnosis of HIV infection in MSM in Spain: greater commitment and action is required. *Enferm. Infecc. Microbiol. Clin.* 36, 463–464. doi: 10.1016/j.eimc.2018.06.003

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

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Flampouris, A., Alvarez, M., Kostaki, V., Chueca, N., Monge, S., García-Bujalance, S., et al. (2014). "Molecular epidemiology analysis of HIV-1 epidemic in Spain," in *VI Congreso National de GeSIDA*, ed. B. A. Grajera (Malaga: Elsevier).

Foley, B. T., Leitner, T., Paraskevis, D., and Peeters, M. (2016). Primate immunodeficiency virus classification and nomenclature: review. *Infect. Genet. Evol.* 46, 150–158. doi: 10.1016/j.meegid.2016.10.018

Giuliani, M., Santoro, M. M., Lo Presti, A., Cella, E., Scognamiglio, P., Lai, A., et al. (2013). Circulation of HIV-1 CRF02-AG among MSM population in Central Italy: a molecular epidemiology-based study. *Biomed. Res. Int.* 2013:810617. doi: 10.1155/2013/810617

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* 30, 1229–1235. doi: 10.1093/molbev/mst012

Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2018). Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* doi: 10.1093/bib/bbx108 [Epub ahead of print].

Kouyos, R. D., von Wyl, V., Yerly, S., Böni, J., Taffé, P., Shah, C., et al. (2010). Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J. Infect. Dis.* 201, 1488–1497. doi: 10.1086/651951

Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Leigh Brown, A. J. (2008). Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 5:e50. doi: 10.1371/journal.pmed.0050050

Mir, D., Jung, M., Delatorre, E., Vidal, N., Peeters, M., and Bello, G. (2016). Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. *Infect. Genet. Evol.* 46, 190–199. doi: 10.1016/j.meegid.2016.05.017

Morán Arribas, M., Rivero, A., Fernández, E., Poveda, T., and Caylá, J. A. (2018). Burden of HIV infection, vulnerable populations and access barriers to healthcare. *Enferm. Infecc. Microbiol. Clin.* 36(Suppl. 1), 3–9. doi: 10.1016/S0213-005X(18)30239-8

Paraskevis, D., Beloukas, A., Stasinos, K., Pantazis, N., de Mendoza, C., Bannert, N., et al. (2019). HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. *BMC Med.* 17:4. doi: 10.1186/s12916-018-1241-1

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tamalet, C., Ravaux, I., Moreau, J., Bregigeon, S., Tourres, C., Richet, H., et al. (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.* 87, 1327–1333. doi: 10.1002/jmv.24184

Von Wyl, V., Kouyos, R. D., Yerly, S., Böni, J., Shah, C., Bürgisser, P., et al. (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

Vourli, G., Pharris, A., Cazein, F., Costagliola, D., Dabis, F., Del Amo, J., et al. (2019). Are European HIV cohort data within EuroCoord representative of the diagnosed HIV population? *Aids* 33, 133–143. doi: 10.1097/QAD.0000000000002034

Yebra, G., De Mulder, M., Martín, L., Rodríguez, C., Labarga, P., Viciana, I., et al. (2012). Most HIV type 1 non-B infections in the spanish cohort of antiretroviral treatment-naïve HIV-infected Patients (CoRIS) are due to recombinant viruses. *J. Clin. Microbiol.* 50, 407–413. doi: 10.1128/JCM.05798-11

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 March 2019
doi: 10.3389/fmicb.2019.00613

Earlier Initiation of Antiretroviral Treatment Coincides With an Initial Control of the HIV-1 Sub-Subtype F1 Outbreak Among Men-Having-Sex-With-Men in Flanders, Belgium

*Lore Vinken[1], Katrien Fransen[2], Lize Cuypers[1], Ivailo Alexiev[3], Claudia Balotta[4], Laurent Debaisieux[5], Carole Seguin-Devaux[6], Sergio García Ribas[2], Perpétua Gomes[7,8], Francesca Incardona[9], Rolf Kaiser[10], Jean Ruelle[11], Murat Sayan[12,13], Simona Paraschiv[14], Roger Paredes[15], Martine Peeters[16], Anders Sönnerborg[17], Ellen Vancutsem[18], Anne-Mieke Vandamme[1,19], Sigi Van den Wijngaert[20], Marc Van Ranst[1,21], Chris Verhofstede[22], Tanja Stadler[23,24], Philippe Lemey[1] and Kristel Van Laethem[1,21]**

[1] Laboratory of Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium, [2] AIDS Reference Laboratory, Department of Clinical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium, [3] National Reference Confirmatory Laboratory of HIV, National Center of Infectious and Parasitic Diseases, Sofia, Bulgaria, [4] Infectious Diseases and Immunopathology Section, 'L. Sacco' Department of Biomedical and Clinical Sciences, 'L. Sacco' Hospital, University of Milan, Milan, Italy, [5] AIDS Reference Laboratory, CUB-Hopital Erasme, Université Libre de Bruxelles, Brussels, Belgium, [6] Laboratory of Retrovirology, Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg, [7] Serviço de Patologia Clínica, Laboratorio de Biologia Molecular, LMCBM, Centro Hospitalar Lisboa Ocidental, Hospital Egas Moniz, Lisbon, Portugal, [8] Centro de Investigação Interdisciplinar Egas Moniz, Instituto Universitário Egas Moniz, Almada, Portugal, [9] EuResist Network GEIE, Rome, Italy, [10] Institute of Virology, University of Cologne, Cologne, Germany, [11] Unit of Medical Microbiology, Institute of Experimental and Clinical Research, Université catholique de Louvain, Brussels, Belgium, [12] PCR Unit, Clinical Laboratory, Kocaeli University, İzmit, Turkey, [13] Research Center of Experimental Health Sciences, Near East University, Nicosia, Cyprus, [14] Molecular Diagnostics Laboratory, National Institute for Infectious Diseases 'Matei Bals', Bucharest, Romania, [15] IrsiCaixa AIDS Research Institute, Universitat Autònoma de Barcelona, Badalona, Spain, [16] UMI 233 TransVIHMI/INSERM1175, Institut de Recherche pour le Développement, University of Montpellier, Montpellier, France, [17] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden, [18] AIDS Reference Laboratory, Department of Microbiology and Infection Control, Universitair Ziekenhuis Brussel, Brussels, Belgium, [19] Unidade de Microbiologia, Center for Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal, [20] AIDS Reference Laboratory, Department of Microbiology, Saint-Pierre University Hospital, Brussels, Belgium, [21] AIDS Reference Laboratory, University Hospitals Leuven, Leuven, Belgium, [22] AIDS Reference Laboratory, Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium, [23] Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland, [24] Swiss Institute of Bioinformatics, Lausanne, Switzerland

OPEN ACCESS

Edited by:
Joris Hemelaar, University of Oxford, United Kingdom

Reviewed by:
Fernando Gonzalez-Candelas, University of Valencia, Spain
Dimitrios Paraskevis, National and Kapodistrian University of Athens, Greece

***Correspondence:**
Kristel Van Laethem
kristel.vanlaethem@uzleuven.be

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *13 October 2018*
Accepted: *11 March 2019*
Published: *26 March 2019*

Citation:
Vinken L, Fransen K, Cuypers L, Alexiev I, Balotta C, Debaisieux L, Seguin-Devaux C, García Ribas S, Gomes P, Incardona F, Kaiser R, Ruelle J, Sayan M, Paraschiv S, Paredes R, Peeters M, Sönnerborg A, Vancutsem E, Vandamme A-M, Van den Wijngaert S, Van Ranst M, Verhofstede C, Stadler T, Lemey P and Van Laethem K (2019) Earlier Initiation of Antiretroviral Treatment Coincides With an Initial Control of the HIV-1 Sub-Subtype F1 Outbreak Among Men-Having-Sex-With-Men in Flanders, Belgium. Front. Microbiol. 10:613. doi: 10.3389/fmicb.2019.00613

Human immunodeficiency virus type 1 (HIV-1) non-B subtype infections occurred in Belgium since the 1980s, mainly amongst migrants and heterosexuals, whereas subtype B predominated in men-having-sex-with-men (MSM). In the last decade, the diagnosis of F1 sub-subtype in particular has increased substantially, which prompted us to perform a detailed reconstruction of its epidemiological history. To this purpose, the Belgian AIDS Reference Laboratories collected HIV-1 *pol* sequences from all sub-subtype F1-infected patients for whom genotypic drug resistance testing was requested

as part of routine clinical follow-up. This data was complemented with HIV-1 *pol* sequences from countries with a high burden of F1 infections or a potential role in the global origin of sub-subtype F1. The molecular epidemiology of the Belgian subtype F1 epidemic was investigated using Bayesian phylogenetic inference and transmission dynamics were characterized based on birth-death models. F1 sequences were retained from 297 patients diagnosed and linked to care in Belgium between 1988 and 2015. Phylogenetic inference indicated that among the 297 Belgian F1 sequences, 191 belonged to a monophyletic group that mainly contained sequences from people likely infected in Belgium (OR 26.67, 95% CI 9.59–74.15), diagnosed in Flanders (OR 7.28, 95% CI 4.23–12.53), diagnosed at a recent stage of infection (OR 7.19, 95% CI 2.88-17.95) or declared to be MSM (OR 34.8, 95% CI 16.0–75.6). Together with a Spanish clade, this Belgian clade was embedded in the genetic diversity of Brazilian subtype F1 strains and most probably emerged after one or only a few migration events from Brazil to the European continent before 2002. The origin of the Belgian outbreak was dated back to 2002 (95% higher posterior density 2000–2004) and birth-death models suggested that its extensive growth had been controlled ($R_e < 1$) by 2012, coinciding with a time period where delay in antiretroviral treatment initiation substantially declined. In conclusion, phylogenetic reconstruction of the Belgian HIV-1 sub-subtype F1 epidemic illustrates the introduction and substantial dissemination of viral strains in a geographically restricted risk group that was most likely controlled by effective treatment as prevention.

Keywords: HIV-1, sub-subtype F1, Belgium, phylodynamics, men-having-sex-with-men, treatment as prevention

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) originated after at least four separate cross-species transmission events from non-human primates to humans in central Africa, which resulted in four groups (M, N, O, and P) that emerged in the early twentieth century (Peeters et al., 2014). Since then, HIV-1 group M lineages have diversified and the current genetic diversity existing within the pandemic group M is the result of subsequent evolution and spread within the human population. Using phylogenetic analysis, HIV-1 group M can be classified into nine major subtypes (A–D, F–H, J, and K) and a few sub-subtypes (like F1 and F2) which reflects a combination of several founder effects and incomplete sampling of the original source population in central Africa (Rambaut et al., 2004). On the other hand, the various circulating recombinant forms (CRFs) originated from co-circulation of subtypes that previously evolved in compartmentalized epidemics.

The HIV-1 subtype B lineage that is responsible for the majority of HIV-1 infections in Europe, was first exported from central Africa to the Caribbean region around 1967 from where it was subsequently introduced in the United States and spread over to other American and European countries, mainly via networks of men-having-sex-with-men (MSM) and people who injects drugs (PWID) (Abecasis et al., 2013; Junqueira and Almeida, 2016; Magiorkinis et al., 2016; Worobey et al., 2016). HIV-1 subtype B was the predominant subtype among Belgians who were diagnosed with HIV-1 between 1985 and 1994 in Antwerp (Belgium), although the presence of other pure subtypes such as A, C, D, F, G, and H already revealed an early introduction of HIV-1 non-B subtypes to the local population (Fransen et al., 1996). Nevertheless, Belgian MSM remained almost exclusively infected with subtype B strains. These findings were confirmed in an independent study where heterosexual contact, African origin and more recent diagnosis were predictors for non-B infection in HIV-1 patients diagnosed between 1983 and 2001 in two other Belgian cities (Brussels and Leuven) (Snoeck et al., 2004). Subsequent local and national surveillance studies that included HIV-1 diagnoses between 1998 and 2012 confirmed the association between subtype B infections, MSM and Caucasian origin. In contrast, HIV-1 non-B infections in Belgium, of which HIV-1 subtype A, C, and CRF02_AG infections were the most prevalent, were associated to migrants from sub-Saharan Africa and individuals showing heterosexual risk behavior (Vercauteren et al., 2008; Chalmet et al., 2010; Pineda-Peña et al., 2014).

In the last decade, sub-subtype F1 diagnoses have markedly increased among newly diagnosed HIV-1 patients linked to care in Belgium. In the light of two Spanish studies that reported lower therapy response rates in HIV-1 sub-subtype F1 infections (Pernas et al., 2014; Cid-Silva et al., 2018), we aimed to investigate the evolutionary history and distinctive socio-demographic and epidemiological features of the HIV-1 sub-subtype F1 epidemic in Belgium. The spatiotemporal origin of the ancestor of the Belgian F1 epidemic and evolutionary links among locations were determined, risk factors were identified, and transmission dynamics were quantified.

MATERIALS AND METHODS

Viral Sequences and Associated Information

Human immunodeficiency virus type 1 sub-subtype F1 *pol* sequences spanning the gene fragments that encode protease (PR) and the 5′-end of reverse transcriptase (RT) were collected from all patients for whom drug resistance testing was requested as part of routine clinical follow-up at six AIDS reference laboratories (ARL) in Antwerp, Brussels, Ghent, and Leuven (up until March 2015). These ARL perform approximately 90% of the HIV-1 drug resistance testing in Belgium. HIV-1 sub-subtype F1 *pol* sequences were requested from partners in other European and African countries where circulation of F1 infections has been reported (Palma et al., 2007; Paraschiv et al., 2009; Frentz et al., 2010; Lai et al., 2012; Ivanov et al., 2013; Sayan et al., 2013; Alexiev et al., 2015). Furthermore, all available HIV-1 sub-subtype F1 *pol* sequences were extracted from the Los Alamos HIV Sequence Database (LANL)[1] (Triques et al., 1999; Montano et al., 2005; Aulicino et al., 2007). Most viral genetic sequences were annotated with sampling time and location, which were used as calibration and discrete traits, respectively, in an analysis of spatiotemporal dispersion patterns (**Supplementary Table S1**). Patient socio-demographics and detailed sample information were collected from mandatory reporting and medical records at the participating ARL in Belgium (gender, year of birth, HIV-1 diagnosis date, infection stage at diagnosis, country of birth, probable country of infection, sexual risk factor for HIV-1 acquisition, viral load, CD4 and CD8 count associated to the HIV-1 *pol* sequence, and therapy initiation date). Infection stage at diagnosis was defined based upon laboratory data: HIV-1 confirmation results with a positive p24 test combined with a negative or undetermined blot or INNO-LIA HIV I/II (Fujirebio Europe, Ghent, Belgium) were classified as acute infections while results with a positive blot lacking a p31 band as recent infections (Verhofstede et al., 2017). In all other instances, at least when records were available, infections were classified as chronic. The study was approved by the Ethical Committee Research UZ/KU Leuven (reference S58359). This study did not require written informed consent by study participants as it only used data collected within routine clinical practice, but patients who opted out of research studies were not included. The entire dataset for this manuscript is not publicly available because it represents a dense sample of our epidemic and inappropriate use could endanger the privacy of the patients. A proportional down-sampling of the dataset (30%) is accessible via GenBank (accession numbers, MK458435-MK458523). Requests to access the entire dataset combined with well-defined project proposals can be mailed to the corresponding author who will submit it to the local ethical committee for ethical and legal clearance when not in conflict with already on-going research.

[1] http://www.hiv.lanl.gov

Quality Control and Compilation of Sequences

The quality of HIV-1 sequences that were generated by the Belgian ARL was assured by quality management systems that are conform to the ISO 15189 standard. In addition, the quality of all collected HIV-1 sequences was subsequently verified using the Quality Control tool hosted on the HIV sequence database LANL. Only sequences with no evidence of hyper-mutations and a maximum of 7 stop codons and/or frameshifts were retained for further analyses. Only pure sub-subtype F1 strains were retained in the study. Therefore, HIV-1 subtype classification was confirmed using the online automated subtyping tools COMET v1.0 (Struck et al., 2014) and Rega v3 (Pineda-Peña et al., 2013). In case of a non-F1 classification by COMET v1.0 or Rega v3 (7% discordant classifications), sequences were subsequently subjected to jpHMM (Schultz et al., 2009; version March 2015). If scored as F1 or F2 sub-subtype by jpHMM, recombination was ruled out using Simplot (Lole et al., 1999). Four HIV-1 subtype B reference sequences were added to the dataset, serving as an outgroup for subsequent phylogenetic analyses (GenBank accession numbers K03455, AY423387, AY173951, and AY331295). Nucleotide sequences were aligned using Muscle v3.7 (Edgar, 2004) and manually edited in MEGA 5 (Tamura et al., 2011). Nucleotide positions encoding for 43 surveillance drug resistance mutations (Bennett et al., 2009) were removed from the alignment to avoid the effect of drug-induced convergent evolution. Duplicate sequences were identified using the ElimDupes tool (LANL) and were removed when originating from the same sample. For patients with multiple sequences, only the first sequence was retained. The final alignment contained 1866 HIV-1 sub-subtype F1 sequences trimmed to the common length of 798 nucleotide positions, spanning sites 2253 to 2549 and 2661 to 3290 (as compared to the HIV-1 subtype B reference sequence HXB2, accession number K03455) (**Supplementary Table S1**).

Phylogenetic and Phylogeographic Analyses

The spatiotemporal origin of the Belgian HIV-1 sub-subtype F1 epidemic was investigated using a Bayesian approach. Prior to this analysis, we examined the accumulation of divergence over the sampling time span using Tempest v1.5 (Rambaut et al., 2016) to assure that sampling dates could be used as calibration. Markov chain Monte Carlo analyses were performed using BEAST v1.8.4 in conjunction with BEAGLE to improve the computational performance (Ayres et al., 2012; Suchard et al., 2018). We specified a GTR nucleotide substitution model with discrete gamma distributed rate variation among sites. Among-lineage variation in evolutionary rates was modeled using a lognormal relaxed molecular clock and a flexible non-parametric Bayesian Skygrid was specified as a coalescent prior. The analyses were run in triplicate for 100 million states and trees were sampled every 10,000th state. Convergence and mixing properties of the Markov chain Monte Carlo chains were assessed with Tracer v1.6 (Rambaut et al., 2018). Results of several chains were combined after removal of the burn-in, and a MCC tree was summarized

using TreeAnnotator within the BEAST package. The timed migration process was reconstructed using a model that allows for non-reversible migration rates. Trees were visualized with FigTree v1.4.3 and iTOL v4.2.3 (Letunic and Bork, 2016).

Phylodynamic Analysis of the Infection Cluster Including Belgian Sequences

Phylodynamic analyses were performed on an infection cluster that mainly contained sequences from our Belgian cohort. A birth-death model was applied to the Belgian sequences included in this infection cluster (191 sequences) in order to quantify transmission dynamics through time. The analyses were performed in BEAST v2 (Bouckaert et al., 2014) using the BDSKY add-on (Stadler et al., 2013). The molecular evolution and clock model were chosen as in the phylogeographic analysis. For the birth-death-skyline parameters, we used a LogNorm(0,1) prior for the effective reproductive number, a LogNorm(0,1) prior for the becoming-non-infectious rate in units per year (i.e., the inverse of the time duration of being infectious in unit of years), and a Unif(0,infinity) prior for the origin of the epidemic. The dimension of the effective reproductive number and the becoming-non-infectious rate was set to 10, meaning the parameters were estimated for 10 equally spaced intervals between the origin and the most recent sample. The sampling proportion was fixed to 0.65 (and 0.8 in a second analysis) for the time between the first and last sample, and 0 prior to the first sample.

Statistical Analysis

Socio-demographic, virological and clinical variables were reported as absolute numbers, percentages and medians with their respective interquartile ranges (IQR), where appropriate. Their association with a Belgian F1 outbreak was evaluated using univariate and multivariate logistic regression in the statistical R software v3.5.1, with statistical significance set at p-value < 0.05. For these analyses, only patients with available information for a specific variable were included.

RESULTS

Increase of Subtype F1 Diagnoses in Belgium

Up to February 2015, 297 HIV-1 sub-subtype F1 infections were newly diagnosed in the participating Belgian ARL (**Figure 1**), of which 137 were analyzed in Antwerp (46%), 95 in Brussels (32%), 38 in Leuven (13%), and 27 in Ghent (9%). The median age of people diagnosed with an HIV-1 sub-subtype F1 infection was 34 years (IQR 29–41) and 77% were men. Over the course of 15 years, the number of HIV-1 sub-subtype F1 diagnoses increased from 3 in March 2001 – February 2003 to 88 in March 2013 – February 2015 (**Figure 1A**). Sub-subtype F1 infections were initially mainly diagnosed in women (65%), but after February 2005 men became the most affected gender (87%). Of the sub-subtype F1 infections for which the respective information was registered, the majority were diagnosed in people born in Belgium (59%, 141/238) and in people who declared that they probably acquired their infection in Belgium (82%, 138/168). The sub-subtype F1 infections were found in homosexual (51%, 117/231), heterosexual (35%, 80/231), and bisexual risk (15%, 34/231) groups (**Figure 1B**).

The Belgian Sub-Subtype F1 Epidemic Within a Global Context

Bayesian phylogenetic analyses were used to estimate the geographical and temporal origin of the HIV-1 sub-subtype F1 strains in our national cohort and to investigate viral dissemination from Belgium to other countries. The final dataset included 1866 sequences that were mainly sampled in Europe (75%) and Brazil (20%) (**Supplementary Table S1**). We found sufficient temporal signal to pursue phylodynamic analysis and used the resulting empirical distribution of the dated phylogenetic trees in a subsequently phylogeographic analysis.

The phylogeographic analysis revealed the divergence of sub-subtype F1 into two well-supported clades for which (i) one clade included mainly Brazilian and Portuguese strains together with a Belgian subclade, while (ii) the other predominantly Belgian, Romanian and Portuguese sequences (upper and lower clade in **Figure 2**, **Supplementary Figure S1**). Their divergence dates back 50 years ago (1968, 95% higher posterior density (HPD) interval of 1963–1973), with an estimated start of local dissemination of HIV-1 F1 in Brazil around 1973 (95% HPD 1969–1975) and subsequently in Portugal around 1979 (95% HPD 1976–1981) (**Figure 2** upper clade). In this upper clade (i), one large Belgian cluster, subsequently referred to as the sub-subtype F1 Belgian outbreak, is embedded within a Western-European clade that in turn is nested within the larger Brazilian clade. The Belgian clade contains four sequences sampled outside Belgium (one from Brazil, one from Luxembourg, and two from Germany), while the Western-European clade consists of the Belgian clade together with three basal sequences from Switzerland, one large well-supported Spanish clade and one Romanian sequence that branches outside the Spanish and Belgian clade. The founder of the HIV-1 sub-subtype F1 strains that are currently circulating in Western-Europe, probably migrated from Brazil to Europe between 1990 and 1999 (95% HPD 1987–2002), whereas the origin of the Belgian and Spanish outbreaks were dated back to 2002 (95% HPD 2000–2004) and 2006 (95% HPD 2004–2008), respectively. Brazil is estimated as the most likely geographical origin of the Western-European clade because the Western-European clade was embedded within the large Brazilian sub-subtype F1 diversity and because one strain that was sampled in Brazil and three strains that were sampled from Brazilians residing in Belgium were also positioned within the Belgian clade. Although Belgium was assigned as ancestral location state of this Western-European clade, the long interconnecting branch suggests that current sampling is insufficient to provide conclusive results as to the specific migration event(s) that finally led to the two large outbreaks in Western Europe (Belgium and Spain).

The remaining sub-subtype F1 sequences from our Belgian cohort represent single lineages or small clusters that are

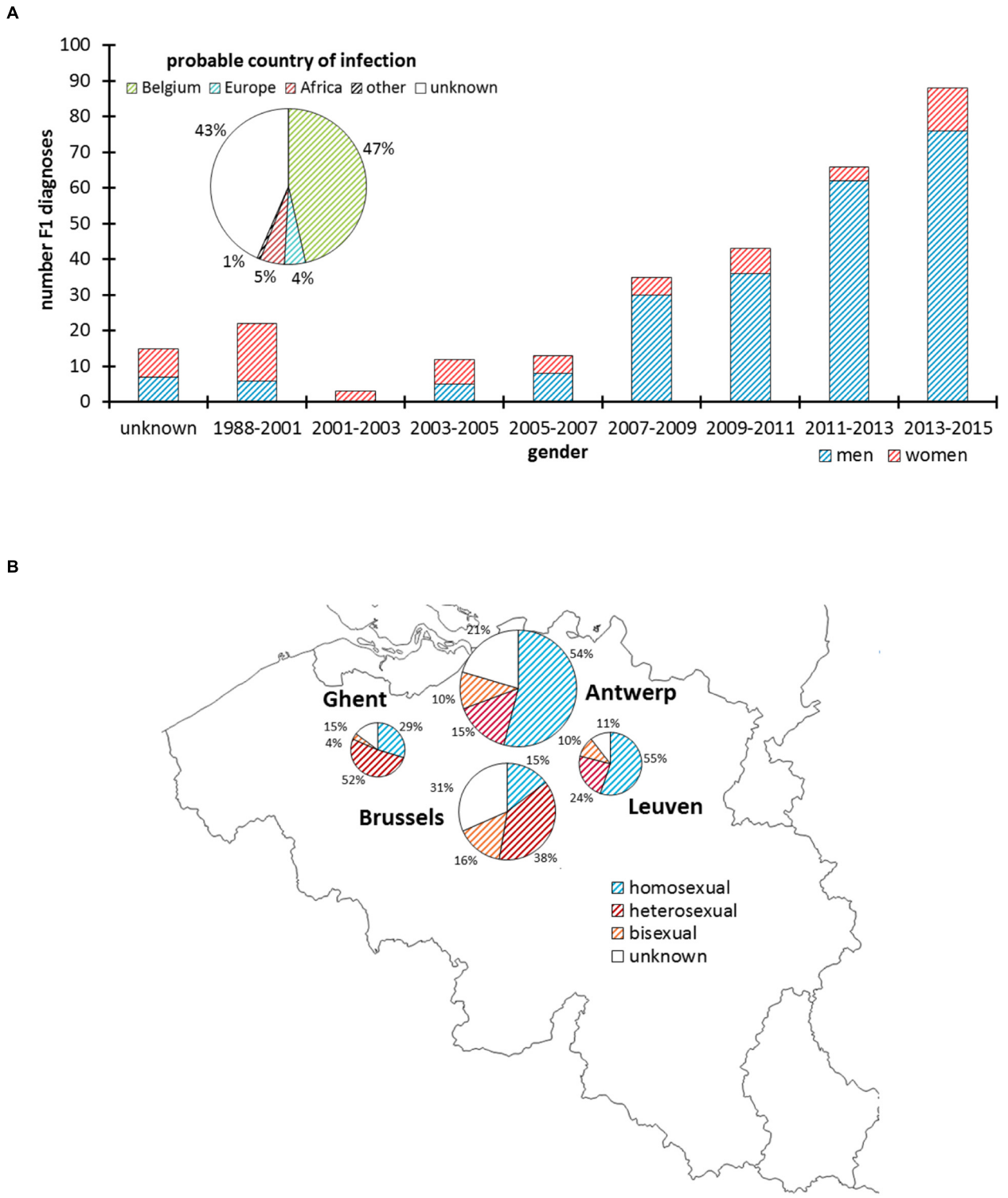


FIGURE 1 | HIV-1 sub-subtype F1 diagnoses in Belgium. Panel **(A)** Number of men and women diagnosed with an HIV-1 sub-subtype F1 infection according to 2-year intervals and based upon a *pol*-genotypic drug resistance test that was requested as part of routine clinical follow-up. The pie chart displays the proportion of people who probably acquired the HIV-1 sub-subtype F1 infection in Belgium, Europe, Africa, or another or unknown region. Panel **(B)** Geographic and proportional distribution of HIV-1 sub-subtype F1 diagnoses in the AIDS Reference Laboratories of Antwerp, Brussels, Leuven, and Ghent. The pie charts display the proportion of people reporting homosexual, heterosexual, or bisexual contacts as risk factor for HIV-1 infection.

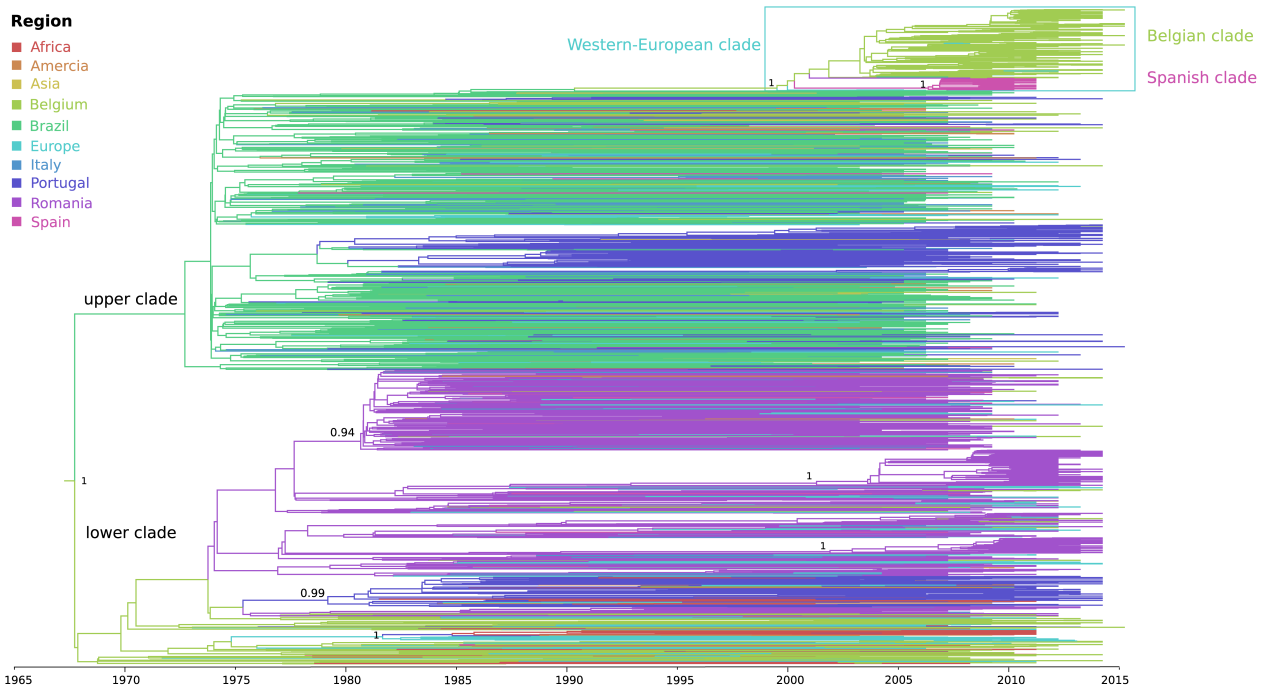


FIGURE 2 | Phylogeographic reconstruction of the Belgian HIV-1 sub-subtype F1 epidemic. The branches in the time-scaled maximum clade credibility tree constructed from the complete dataset of 1866 *pol* sequences are colored by their most probable location state (see legend), which could be a specific sampling country or grouped as a continent when less than 30 sequences were available for a specific country (**Supplementary Table S1**). The HIV-1 sub-subtype F1 strains that are associated with the local outbreak (191 lineages), most probably entered Belgium via Latin-America, more particularly Brazil, a transmission event that is estimated to have occurred around 2002 (95% HPD 2000 – 2004). The 106 Belgian sequences that do not belong to the monophyletic group, are dispersed within less recent clades, grouping with lineages from Africa, Latin-America and Europe. Posterior root node support is visualized in a selection of nodes.

dispersed across the tree and are intermixed among viral lineages from Africa, Latin-America and Europe (**Figures 2**, **3**). Several other strains from Sub-Saharan migrants residing in Belgium but who originated from the Democratic Republic of Congo (DRC), Congo, and Angola occupied the most basal positions of the tree (**Figure 2** lower clade, **Figure 3**). They cluster with African lineages that probably migrated to Europe between 1972 and 1977 (95% HPD 1969–1980) and that were at the origin of several larger sub-subtype F1 outbreaks in Europe, more specifically in Portugal (1979, 95% HPD 1976–1982), Romania (1981, 95% HPD 1978–1983; 2001, 95% HPD 1996–2006; and 2002, 95% HPD 1998–2006), and Bulgaria (1982, HPD 1979–1985). As, we defined a location trait according to country of sampling in our analysis, the over-representation of Belgian sequences – probably to a large extent reflecting African diversity – led Belgium to be inappropriately assigned as the ancestor location of this lower clade that contained many African, Romanian, Portuguese, and other European lineages (**Figure 2** lower clade).

Socio-Epidemiological Correlates With the Sub-Subtype F1 Outbreak in Belgium

As one or only a few sub-subtype F1 lineages have successfully established local transmissions among 191 of 297 people diagnosed with a sub-subtype F1 infection in Belgium (64%), we used univariate and multivariate logistic regression to identify socio-epidemiological variables that were associated with this successful sub-subtype F1 outbreak in Belgium (**Table 1**). HIV-1 patients who were diagnosed with a sub-subtype F1 infection at the AIDS reference laboratories in Antwerp (OR 6.80, 95% CI 3.86–11.99), Leuven (OR 4.22, 95% CI 1.59–11.16) or more broadly in the region of Flanders (OR 7.28, 95% CI 4.23–12.53) had significantly higher odds of belonging to this F1 outbreak, compared to people who were diagnosed with a sub-subtype F1 infection elsewhere in Belgium. In addition, many other variables of the sub-subtype F1 infections diagnosed in Belgium displayed significantly higher odds of being part of the outbreak, of which the most informative were male gender (OR 52.42, 95% CI 19.90–138.08), infection in MSM (combining homosexuals and bisexuals) (OR 34.8, 95% CI 16.0–75.6), probable infection in Belgium (OR 26.67, 95% CI 9.59–74.15), born in Belgium (OR 12.10, 95% CI 6.35-23.08), recent infection stage at diagnosis (OR 7.19, 95% CI 2.88–17.95) and higher viral load (OR 2.36, 95% CI 1.78–3.13). In multivariate analysis, sub-subtype F1 infections in MSM and in people born in Belgium remained to display a higher likelihood for belonging to the sub-subtype F1 Belgian outbreak [OR 7.82, 95% CI 3.80–16.11 and OR 2.01 (1.06–3.78), respectively]. The variables sampling region, ARL in Belgium, gender, country of birth and sexual risk group were visualized for all Belgian sequences in a circular representation of the phylogenetic tree (**Figure 3**).

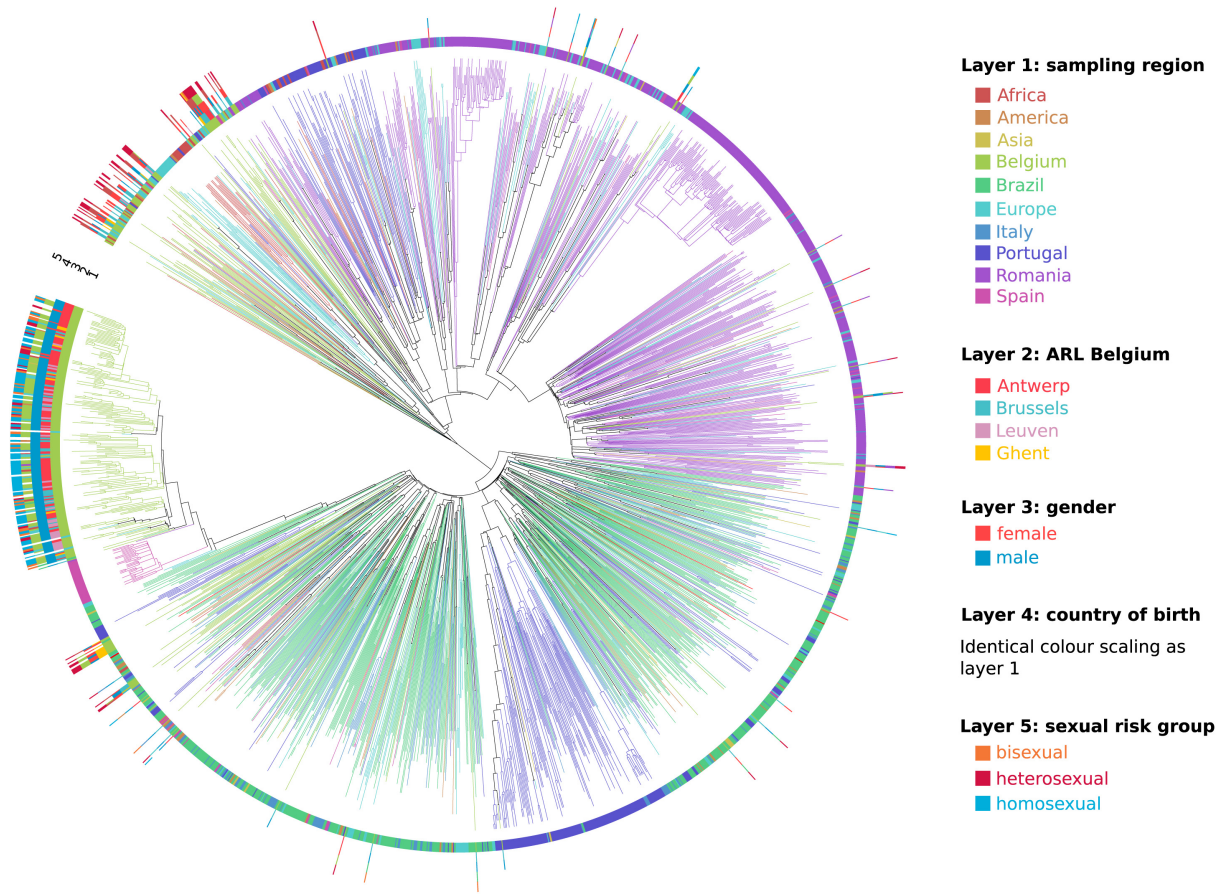


FIGURE 3 | Circular representation of the maximum clade credibility tree of the Belgian sub-subtype F1 epidemic within a global context. Branches in the tree are colored according to sampling region (identical coloring scaling as in **Figure 2**), which is also displayed in layer 1. For the HIV-1 sub-subtype F1 strains that were sampled at the AIDS reference laboratories (ARL) in Belgium, the location of testing was displayed in layer 2, gender in layer 3, country of birth in layer 4, and sexual risk group in layer 5. This detailed information was largely lacking for the sequences sampled outside Belgium.

Reduced Onward Transmission of Sub-Subtype F1 Strains Coincident With an Earlier Initiation of Antiretroviral Therapy

Since the onset of the sub-subtype F1 outbreak in Belgium (2002, 95% HPD 2000–2004), we covered a time period where antiretroviral therapy was initially still deferred in asymptomatic patients with CD4 count above 350 cells/mm^3 (European AIDS Clinical Society, 2005). This general practice gradually switched to the consideration of antiretroviral therapy for asymptomatic patients with CD4 counts between 350 and 500 cells/mm^3 (European AIDS Clinical Society, 2007) and above 500 cells/mm^3 (European AIDS Clinical Society, 2013). Given the benefits of antiretroviral therapy in reducing infectivity, we wanted to quantify transmission dynamics through time using a birth-death model and monitor the trends in therapy uptake in our specific setting. Therapy and diagnosis information was available for 147 HIV-1 patients whose viral strains belonged to the sub-subtype F1 outbreak in Belgium (77%, 147/191). The delay in therapy initiation decreased from a median of 1011 days (IQR 368–1683) or almost 3 years for patients diagnosed in 2005 and 2006 to a median of 61 days (IQR 26–208) or about 2 months for patients diagnosed in 2013 and 2014 (**Figure 4**). In the birth-death model, assuming a constant 65% sampling proportion, the number of expected secondary infections per infected person (effective reproductive number R_e) was estimated to be a median of 3 in 2005 and gradually declined to a median of 1.5 in 2011, after which a substantial drop to below 1 was observed (**Figure 5**). During this time window, the median infectious time also declined and the number of expected secondary infections per infected person and per year (transmission rate) decreased from a median of 1 to well below 1. Sensitivity analysis using a constant sampling proportion of 80%, as the diagnosis and sequencing rate is expected to be higher in MSM, showed similar results and suggests that the estimates are relatively robust to prior specification on the sampling proportion.

DISCUSSION

Human immunodeficiency virus type 1 subtype F most likely emerged from the early 20th century HIV-1 diversity

TABLE 1 | Characteristics of people diagnosed with an HIV-1 sub-subtype F1 infection in Belgium and factors associated with this local outbreak.

Characteristic	Local outbreak (*n* = 191)	Other (*n* = 106)	Univariate logistic regression		Multivariate logistic regression	
			OR (95% CI)	*p*	OR (95% CI)	*p*
AIDS Reference Laboratory [% (*n*)]						
Antwerp	61.3 (117)	18.9 (20)	6.80 (3.86–11.99)	< 0.0001		
Brussels	16.7 (32)	59.4 (63)	0.14 (0.08–0.24)	< 0.0001		
Leuven	17.3 (33)	4.7 (5)	4.22 (1.59–11.16)	0.0037		
Ghent	4.7 (9)	17.0 (18)	0.24 (0.10–0.56)	0.0009		
Region [% (*n*)]						
Flanders	83.3 (159)	40.6 (43)	7.28 (4.23–12.53)	< 0.0001		
Gender [% (*n*)]						
Male	97.4 (186)	41.5 (44)	52.42 (19.90–138.08)	< 0.0001		
Age at diagnosis [years]						
Median (IQR)	34 [29, 41] (190)	34 [27, 40] (92)	1.03 (1.00–1.06)	0.0343		
Diagnosis date [dd-mm-yyyy]						
Median [IQR] (*n*)	09-05-2012 [21-03-2010, 23-11-2013] (190)	24-1-2008 [3-4-2002, 31-10-2011] (92)	1.33 (1.23–1.45)	< 0.0001		
Infection stage at diagnosis [% (*n*)]						
Acute	10.1 (16/158)	2.2 (1/45)	4.96 (0.64–38.45)	0.1256		
Recent	52.5 (83/158)	13.3 (6/45)	7.19 (2.88–17.95)	< 0.0001		
Chronic	37.4 (59/158)	84.5 (38/45)	0.11 (0.05–0.26)	< 0.0001		
Unknown	17.2 (33)	57.5 (61)				
Country of birth [% (*n*)]						
Belgium	77.8 (123/158)	22.5 (18/80)	12.10 (6.35–23.08)	< 0.0001	2.01 (1.06–3.78)	0.0316
Other	22.2 (35/158)	77.5 (62/80)				
Unknown	17.2 (33)	24.5 (26)				
Probable country of infection [% (*n*)]						
Belgium	95.2 (120/126)	42.9 (18/42)	26.67 (9.59–74.15)	< 0.0001		
Other	4.8 (6/126)	57.1 (24/42)				
Unknown	34.0 (65)	60.4 (64)				
Sexual risk group [% (*n*)]						
Homosexual contacts	69.2 (110/159)	9.7 (7/72)	20.85 (8.92–48.73)	< 0.0001		< 0.0001
Bisexual contacts	18.2 (29/159)	7.0 (5/72)	2.99 (1.11–8.08)			
Heterosexual contacts	12.6 (20/159)	83.3 (60/72)	0.03 (0.01–0.06)	0.0308		
Men-having-sex-with-men	87.4 (139/159)	16.7 (12/72)	34.8 (16.0–75.6)	< 0.0001	7.82 (3.80–16.11)	
Unknown	16.8 (32)	32.1 (34)		< 0.0001		
Viral load at sequencing [log_{10} copies/ml] Median [IQR] (*n*)	5.18 [4.61–5.79] (191)	4.36 [3.75, 4.92] (101)	2.36 (1.78–3.13)	< 0.0001		
CD4+ at sequencing [cells/μl]						
Median [IQR] (n)	488 [358–645] (184)	334 [175, 528] (89)	1.00 (1.00–1.00)	< 0.0001		
CD8+ at sequencing [cells/μl]						
Median [IQR] (*n*)	1059 [697–1598] (173)	1041 [778, 1390] (67)	1.00 (1.00–1.00)	0.2980		

A univariate and multivariate logistic regression analysis was performed to identify socio-epidemiological variables that were associated with the local HIV-1 F1 outbreak in Belgium (n = 191). A p-value of < 0.05 was considered to be significant. Abbreviations: n, number, IQR, interquartile range, OR, odds ratio, CI, confidence interval.

in the DRC, where its current prevalence remains less than 5% (Vidal et al., 2000; Faria et al., 2014; Rodgers et al., 2017). This subtype accounts for less than 1% of HIV-1 infections worldwide and is subdivided into two sub-subtypes F1 and F2, of which the detection of the latter remains mostly restricted to Cameroon (Brennan et al., 2008; Hemelaar et al., 2011). In contrast, sub-subtype F1 strains have spread to other Central African countries, reaching a particularly high prevalence of 14% in Angola (Afonso et al., 2012). In Belgium, the burden of sub-subtype F1 infections among newly diagnosed HIV-1 patients increased to 11.1% in 2015. The aim of this study was to characterize the epidemic dynamics of this non-B subtype epidemic in Belgium by using phylogenetic approaches and the analysis of socio-epidemiological variables.

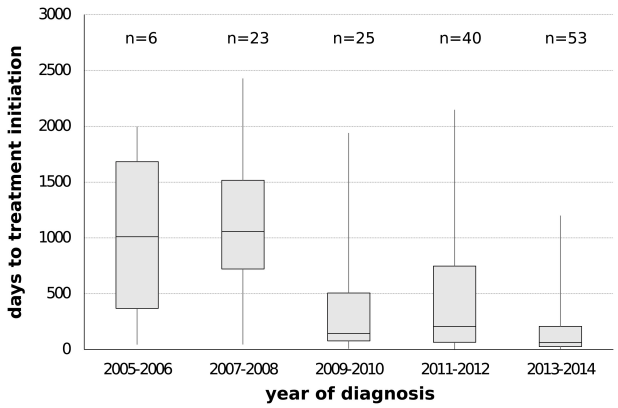


FIGURE 4 | Delay in antiretroviral treatment among people infected with sub-subtype F1 strains belonging to the local outbreak in Belgium. The number of days between diagnosis date and treatment initiation date were plotted according to 2-year intervals of diagnosis date. Information on prescribed antiviral therapy and diagnosis date was available for 77% of the HIV-1 patients belonging to the local outbreak in Belgium (147/191).

The Dissemination of HIV-1 Sub-Subtype F1 Across the World

Our and other phylogenetic analyses confirmed that the most basal F1 sequences were isolated from people originating from Central African countries and that the major migration events out of Africa occurred in the 1970s and early 1980s seeding some larger F1 epidemics over the world (Mehta et al., 2011; Bello et al., 2012; Lai et al., 2014). During the Angolan civil war that started in 1975 and lasted for more than 16 years, an extensive migration of troops, settlers and refugees was observed between Angola, DRC, Portugal, Brazil, Cuba, Russia, and several Eastern-European countries and this is considered as an important factor in the global dissemination of F1. In Brazil, sub-subtype F1 infections have reached prevalence rates of more than 7% and the strains are currently circulating among various risk groups (Brindeiro et al., 2003). It is the most probable epicenter of F1-related epidemics within the Southern Cone and one of the sources of the Portuguese sub-subtype F1 epidemic (Aulicino et al., 2007; Bello et al., 2007, 2012; Carvalho et al., 2015). Unrelated to the Brazilian epidemic, sub-subtype F1 was initially introduced and spread among adult heterosexuals in Romania. In this Eastern-European country, it subsequently reached a very high prevalence rate of more than 70% due to several major outbreaks among children in orphanages and people using intravenous drugs (Dumitrescu et al., 1994; Bandea et al., 1995; Apetrei et al., 1998; Op De Coul et al., 2000; Guimarães et al., 2009; Paraschiv et al., 2009; Mehta et al., 2011; Mbisa et al., 2012; Niculescu et al., 2014). Both Brazil and Romania were at the origin of the F1 epidemic in Italy, where Brazilian strains contributed to many more introduction events characterized with a more extensive spread (Lai et al., 2014). Since then, the prevalence of sub-subtype F1 increased gradually among Italian descendants and mainly among self-declared heterosexual men, although phylogenetic analyses suggested that some of these self-declared heterosexual men might have been infected through homosexual contact (Bracciale et al., 2009; Lai et al., 2012, 2014).

Although our and a previous study indicated that another sub-subtype F1 lineage migrated from Brazil to Europe during the 1990s, this lineage obtained only more recently epidemic proportions in two countries, respectively, about 16 years ago in Belgium and 12 years ago in Spain (Paraskevis et al., 2015). In both instances, F1 was introduced in local MSM networks, a factor probably important in its subsequent spread and circulation at high rates (Thomson et al., 2012). Our analyses did not indicate any intermixing between both epidemics and only sporadic and putative exports from Belgium to Luxembourg (one) and Germany (two) were detected. This is in agreement with another study that traced the HIV-1 subtype B mobility in Europe and that characterized Belgium primarily as a sink of HIV-1 lineages rather than a source (Paraskevis et al., 2009). Although, we confirmed previous findings on the global spread of HIV-1 sub-subtype F1, our sampling approach might not have been sensitive enough to capture the specific migration events that resulted into these two large outbreaks or into extensive onward transmission to other countries. For our analyses, we collected HIV-1 *pol* sequences from countries with a known high burden of F1 infections or a potential role in the global spread of sub-subtype F1. In this dataset, Belgium, Romania, Portugal, Luxembourg, and Bulgaria were the best-represented countries as they contributed a dense sampling, whereas Italy provided a down-sampling proportional to their local sub-subtype F1 diversity. However, other countries might have been undersampled as they were represented by sequences that were received from single cohorts or by sequences that were collected from public databases. Such sample biases call for caution in interpreting the results of phylogeographic estimates.

The Belgian HIV-1 Sub-Subtype F1 Outbreak

The first HIV-1 sub-subtype F1 infections within the local outbreak cluster were diagnosed in 2005. It concerned one chronic infection in Leuven and one infection with unknown status in Antwerp, suggesting a time of about 3 years between the estimated introduction in Belgium (2002, 95% HPD 2000–2004) and the first case reported by the Belgian ARL. Although there is universal access to HIV-1 health care in Belgium and we included all patients for whom genotypic drug resistance testing indicated a sub-subtype F1 infection, whether at baseline or at therapeutic failure, our sequencing efforts in routine practice only cover approximately 60% of all newly HIV-1 diagnoses registered in the national database. However, we believe that we do capture the majority of patients who are contributing to the local spread, as sub-subtype F1 is mainly circulating among MSM, a group who is well engaged in medical care (Van Beckhoven et al., 2015). Any HIV-1 patient who seeks medical care in Belgium will receive a confirmation of his/her infection and will therefore be registered in the national database. Therefore, we believe that a large proportion of the people living with HIV-1 in Belgium who do not receive a baseline resistance test at diagnosis are people with an infection that was already

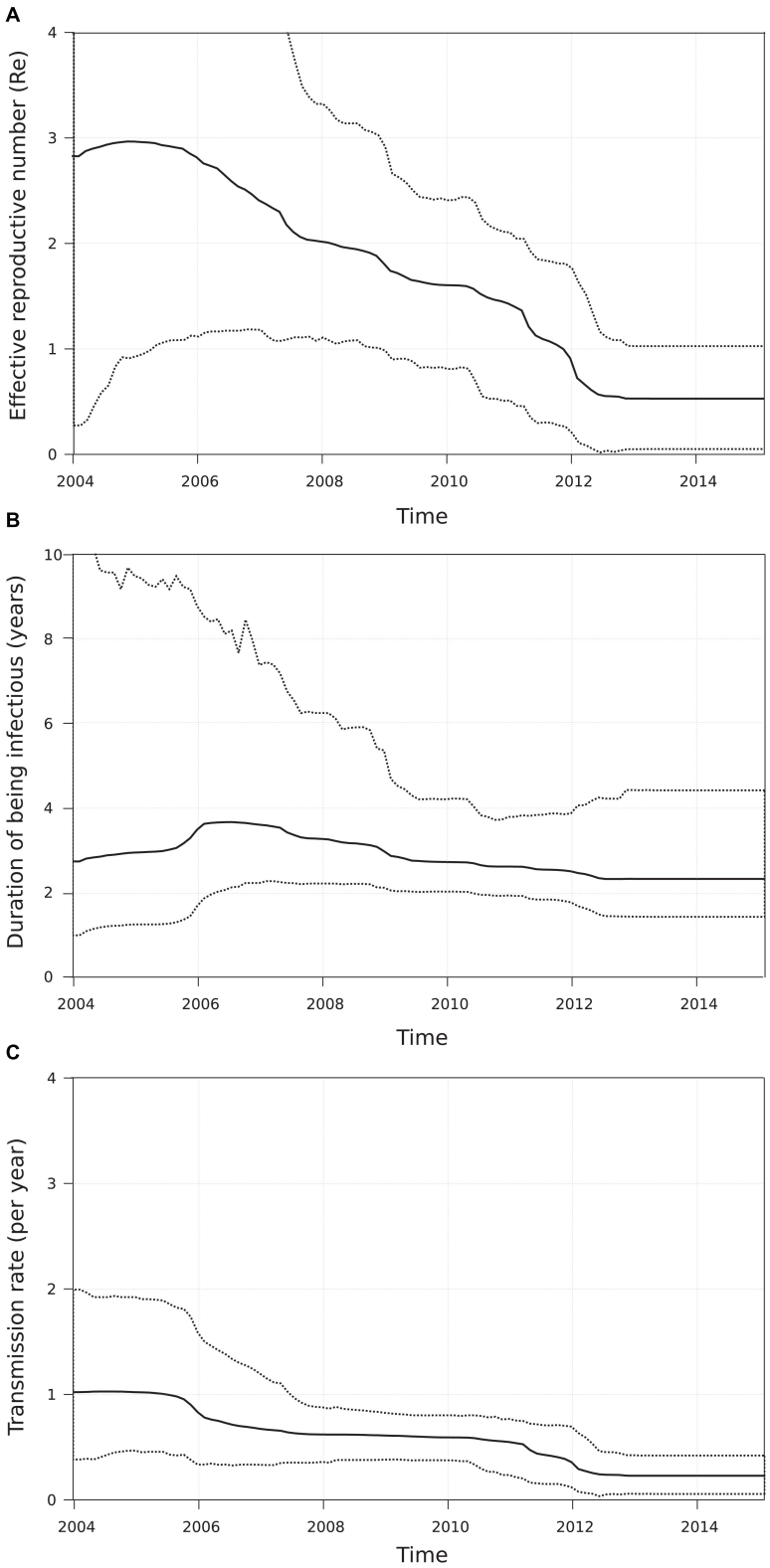


FIGURE 5 | Birth-death model applied to the sub-subtype F1 outbreak in Belgium. For all patients belonging to the large Belgian cluster (n = 191), the number of secondary infections per infected person (effective reproductive number Re, panel **(A)**), the median infectious time in years (panel **(B)**) and the number of secondary infections per infected person per year (transmission rate, panel **(C)**) was estimated. After the onset of the outbreak both Re and transmission rate started to decline, with the largest drop occurring after 2011.

acquired and diagnosed abroad and who are already on successful antiretroviral treatment upon arrival in Belgium, or these people are only temporarily in the country and do not enter the HIV-1 health care system in Belgium.

With the dense sampling of our local epidemic at hand, we aimed to estimate the effective reproductive number R_e, the transmission (birth) rates and the time of being infectious in our viral population using birth-death modeling in order to gain insights into the dynamics of the outbreak. The initial high R_e-value, high transmission rate and long times of being infectious were in agreement with the explosive increase of sub-subtype F1 infections in our national cohort. However, the mean R_e-value declined substantially from 2012 onward and this event was associated with lower transmission rates and shorter durations of being infectious. These observations may be explained by the policy changes around that time when therapy eligibility thresholds for asymptomatic HIV-1 patients became less strict in Belgium. Indeed, medical records showed a remarkable dent in the median treatment delay among people belonging to our outbreak around 2009–2010. Although this relationship does not demonstrate causality, it is supported by studies where viral suppression by antiretroviral therapy has shown to be effective in preventing HIV-1 transmission among serodiscordant heterosexual and MSM couples (Cohen et al., 2011; Rodger et al., 2016; Bavinton et al., 2018). However, the relevant changes in transmission dynamics were not yet reflected in the number of new sub-subtype F1 diagnoses up to the start of 2015. This could suggest that the effects were too recent to be fully captured by our surveillance data of new diagnoses or that the conditions were not yet fully optimal to achieve the full potential of treatment as prevention in this outbreak (Okano et al., 2016). Although a substantial proportion of our outbreak was diagnosed at an acute and recent stage of infection, we should aim to increase testing and diagnosis rates among people at risk of acquiring HIV-1 and treat them without any delay. In this respect, the INSIGHT START study resolved the last concerns for immediate initiation of antiretroviral therapy in asymptomatic HIV-1 infections as this study showed that this strategy was superior to deferral until the CD4 count declined to 350 cells/mm^3 and that it was not associated with an increased rate of adverse events (INSIGHT START Study Group et al., 2015).

In conclusion, this study revealed that sub-subtype F1 strains were introduced in Belgium around 2002 and that during a decade they spread exponentially among MSM who were mainly from Belgian descent, residing in Flanders and diagnosed at a recent stage of infection. Our birth-death phylodynamic estimates are in line with the expectations of the effectiveness of treatment as prevention because we identify a decline in the onward transmission of sub-subtype F1 strains that coincides with the implementation of earlier initiation of antiretroviral therapy in our cohort. Further follow-up is planned, especially within the context of the roll-out of antiretroviral therapy in early asymptomatic HIV-1 infections and of the implementation of pre-exposure prophylaxis which is reimbursed to people who are at increased risk for HIV-1 infection since June 2017.

DATA AVAILABILITY

The datasets generated for this study can be found in GenBank, MK458435-MK458523.

AUTHOR CONTRIBUTIONS

KVL conceived and designed the study. LV, KF, IA, CB, LD, CS-D, SGR, PG, FI, RK, JR, MS, SP, RP, MP, AS, EV, SVW, MVR, CV, A-MV, and KVL collected and analyzed the epidemiological and viral data. LV, LC, TS, and PL performed the phylogenetic analyses and inferences. LV, LC, and KVL wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported in part by grants from the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) (G.0692.14N, G.0E84.16N, G.0B23.17N) and KU Leuven (Program Financing No. PF/10/018), and by the AIDS Reference Laboratories that receive support from the Belgian Ministry of Social Affairs through a fund within the Health Insurance System. The computational resources and services used in this work were partly provided by the Hercules Foundation and the Flemish Government – department EWI-FWO Krediet aan Navorsers (KAN2012 1.5.249.12). PL acknowledges funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 725422-ReservoirDOCS), from the Special Research Fund, KU Leuven ("Bijzonder Onderzoeksfonds," KU Leuven, OT/14/115), and the Research Foundation – Flanders ("Fonds voor Wetenschappelijk Onderzoek – Vlaanderen," G.0662.15N, G.0D51.17N, and G.0B93.17N). TS was supported in part by the European Research Council under the 7th Framework Program of the European Commission (PhyPD, Grant Agreement 335529).

ACKNOWLEDGMENTS

This study was the result of a collaborative effort. We would like to thank Andrea-Clemencia Pineda-Peña, Yoeri Schrooten, Fossie Ferreira, and Tim Dierickx for their technical assistance, and the patients and clinicians at the AIDS Reference Centers who contributed data. We would like to acknowledge BREACH that facilitates HIV research among different actors in Belgium.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00613/full#supplementary-material

REFERENCES

Abecasis, A. B., Wensing, A. M., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A., et al. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 10:7. doi: 10.1186/1742-4690-10-7

Afonso, M. J., Bello, G., Guimarães, M. L., Sojka, M., and Morgado, M. G. (2012). HIV-1 diversity and transmitted drug resistance mutations among patients from the North, Central and South Regions of Angola. *PLoS One* 7:e42996. doi: 10.1371/journal.pone.0042996

Alexiev, I., Shankar, A., Wensing, A. M., Beshkov, D., Elenkov, I., Stoycheva, M., et al. (2015). Low HIV-1 transmitted drug resistance in Bulgaria against a background of high clade diversity. *J. Antimicrob. Chemother.* 70, 1874–1880. doi: 10.1093/jac/dkv011

Apetrei, C., Necula, A., Holm-Hansen, C., Loussert-Ajaka, I., Pandrea, I., Cozmei, C., et al. (1998). HIV-1 diversity in Romania. *AIDS* 12, 1079–1085. doi: 10.1097/00002030-199809000-00016

Aulicino, P. C., Bello, G., Rocco, C., Romero, H., Mangano, A., Morgado, M. G., et al. (2007). Description of the first full-length HIV type 1 subtype F1 strain in Argentina: implications for the origin and dispersion of this subtype in South America. *AIDS Res. Hum. Retroviruses* 23, 1176–1182. doi: 10.1089/aid.2007.0038

Ayres, D. L., Darling, A., Swickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., et al. (2012). BEAGLE : an application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* 61, 170–173. doi: 10.1093/sysbio/syr100

Bandea, C. I., Ramos, A., Pieniazek, D., Pascu, R., Tanuri, A., G, et al. (1995). Epidemiologic and evolutionary relationships between Romanian and Brazilian HIV-subtype F strains. *Emerg. Infect. Dis.* 1, 91–93. doi: 10.3201/eid0103.950305

Bavinton, B. R., Pinto, A. N., Phanuphak, N., Grinsztejn, B., Prestage, G. P., Zablotska-Manos, I. B., et al. (2018). Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. *Lancet HIV* 5, e438–e447. doi: 10.1016/S2352-3018(18)30132-2

Bello, G., Afonso, J. M., and Morgado, M. G. (2012). Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect. Genet. Evol.* 12, 1079–1086. doi: 10.1016/j.meegid.2012.03.014

Bello, G., Eyer-Silva, W. A., Couto-Fernandez, J. C., Guimarães, M. L., Chequer-Fernandez, S. L., Teixiera, S. L., et al. (2007). Demographic history of HIV-1 subtypes B and F in Brazil. *Infect. Genet. Evol.* 7, 263–270. doi: 10.1016/j.meegid.2006.11.002

Bennett, D., Camacho, R. J., Otelea, D., Kuritzkes, D. R., Fleury, H., Kiuchi, M., et al. (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 4:e4724. doi: 10.1371/journal.pone.0004724

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Bracciale, L., Colafigli, M., Zazzi, M., Corsi, P., Meraviglia, P., Micheli, V., et al. (2009). Prevalence of transmitted HIV-1 drug resistance in HIV-1 infected patients in Italy: evolution over 12 years and predictors. *J. Antimicrob. Chem.* 64, 607–615. doi: 10.1093/jac/dkp246

Brennan, C. A., Bodelle, P., Coffey, R., Devare, S. G., Golden, A., Hackett, J. Jr., et al. (2008). The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. *J. Acquir. Immune Defic. Syndr.* 49, 432–439. doi: 10.1097/QAI.0b013e31818a6561

Brindeiro, R. M., Diaz, R. S., Sabino, E. C., Morgado, M. G., Pires, I. L., Brigido, L., et al. (2003). Brazilian network for HIV drug resistance surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS* 17, 1063–1069. doi: 10.1097/00002030-200305020-00016

Carvalho, A., Costa, P., Triunfante, V., Branca, F., Rodrigues, F., Santos, C. L., et al. (2015). Analysis of a local HIV-1 epidemic in Portugal highlights established transmission of non-B and non-G subtypes. *J. Clin. Microbiol.* 53, 1506–1514. doi: 10.1128/JCM.03611-14

Chalmet, K., Staelens, D., Blot, S., Dinakis, S., Pelgrom, J., Plum, J., et al. (2010). Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infect. Dis.* 10:262. doi: 10.1186/1471-2334-10-262

Cid-Silva, P., Margusino-Framiñán, L., Balboa-Barreiro, V., Martín-Herranz, I., Castro-Iglesias, Á, Pernas-Souto, B., et al. (2018). Initial treatment response among HIV subtype F infected patients who started antiretroviral therapy based on integrase inhibitors. *AIDS* 32, 121–125. doi: 10.1097/QAD.0000000000001679

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Cohen, M. S., Chen, Y. Q., et al. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med.* 365, 493–505. doi: 10.1056/NEJMoa1105243

Dumitrescu, O., Kalish, M. L., Kliks, S. C., Bandea, C. I., and Levy, J. A. (1994). Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J. Infect. Dis.* 169, 281–288. doi: 10.1093/infdis/169.2.281

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

European AIDS Clinical Society (2005). *EACS Guidelines. [Internet]*. Available at: http://www.eacsociety.org/guidelines/eacs-guidelines/eacs-guidelines.html

European AIDS Clinical Society (2007). *EACS Guidelines. [Internet]*. Available at: http://www.eacsociety.org/guidelines/eacs-guidelines/eacs-guidelines.html

European AIDS Clinical Society (2013). *EACS Guidelines. [Internet]*. Available at: http://www.eacsociety.org/guidelines/eacs-guidelines/eacs-guidelines.html

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., et al. (2014). HIV epidemiology. The early spread and epidemic ignition of *HIV*-1 in human populations. *Science* 346, 56–61. doi: 10.1126/science.1256739

Fransen, K., Buvé, A., Nkengasong, J. N., Laga, M., and van der Groen, G. (1996). Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. *Lancet* 347:1403. doi: 10.1016/S0140-6736(96)91042-9

Frentz, D., Boucher, C. A., Assel, M., De Luca, A., Fabbiani, M., Incardona, F., et al. (2010). Comparison of HIV-1 genotypic resistance test interpretation systems in predicting virological outcomes over time. *PLoS One* 5:e11505. doi: 10.1371/journal.pone.0011505

Guimarães, M. L., Vicente, A. C., Otsuki, K., da Silva, R. F., Francisco, M., da Silva, F. G., et al. (2009). Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology* 6:39. doi: 10.1186/1742-4690-6-39

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and WHO-UNAIDS Network for HIV Isolation, and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 25, 679–689. doi: 10.1097/QAD.0b013e328342ff93

INSIGHT START Study Group, Lundgren, J. D., Babiker, A. G., Gordin, F., Emery, S., Grund, B., et al. (2015). Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N. Engl. J. Med.* 373, 795–807. doi: 10.1056/NEJMoa1506816

Ivanov, I. A., Beshkov, D., Shankar, A., Hanson, D. L., Paraskevis, D., Georgieva, V., et al. (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

Junqueira, D. M., and Almeida, S. E. (2016). HIV-1 subtype B: traces of a pandemic. *Virology* 495, 173–184. doi: 10.1016/j.virol.2016.05.003

Lai, A., Ciccozzi, M., Franzetti, M., Simonetti, F. R., Bozzi, G., Binda, F., et al. (2014). Local and global spatio-temporal dynamics of HIV-1 subtype F1. *J. Med. Virol.* 86, 186–192. doi: 10.1002/jmv.23783

Lai, A., Simonetti, F. R., Zehender, G., De Luca, A., Micheli, V., Meraviglia, P., 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. doi: 10.1371/journal.pone.0042223

Letunic, I., and Bork, P. (2016). Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Magiorkinis, G., Angelis, K., Mamais, I., Katzourakis, A., Hatzakis, A., Albert, J., et al. (2016). The global spread of HIV-1 subtype B epidemic. *Infect. Genet. Evol.* 46, 169–179. doi: 10.1016/j.meegid.2016.05.041

Mbisa, J. L., Hué, S., Buckton, A. J., Myers, R. E., Duiculescu, D., Ene, L., et al. (2012). Phylodynamic and phylogeographic patterns of the HIV type 1

subtype F1 parenteral epidemic in Romania. *AIDS Res. Hum. Retroviruses* 2012, 1161–1166. doi: 10.1089/AID.2011.0390

Mehta, S. R., Wertheim, J. O., Delport, W., Ene, L., Tardei, G., Duiculescu, D., et al. (2011). Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. *Infect. Genet. Evol.* 11, 975–979. doi: 10.1016/j.meegid.2011.03.009

Montano, S. M., Sanchez, J. L., Laguna-Torres, A., Cuchi, P., Avila, M. M., Weissenbacher, M., et al. (2005). Prevalences, genotypes, and risk factors for HIV transmission in South America. *J. Acquir. Immune Defic. Syndr.* 40, 57–64. doi: 10.1097/01.qai.0000159667.72584.8b

Niculescu, I., Paraschiv, S., Paraskevis, D., Abagiu, A., Batan, I., Banica, L., et al. (2014). Recent HIV-1 outbreak among intravenous drug users in Romania: evidence for cocirculation of CRF14_BG and subtype F1 strains. *AIDS Res. Hum. Retroviruses* 31, 488–495. doi: 10.1089/aid.2014.0189

Okano, J. T., Robbins, D., Palk, L., Gerstoft, J., Obel, N., and Blower, S. (2016). Testing the hypothesis that treatment can eliminate HIV: a nationwide, population-based study of the Danish HIV epidemic in men who have sex with men. *Lancet Infect. Dis.* 16, 789–796. doi: 10.1016/S1473-3099(16)30022-6

Op De Coul, E., van den Burg, R., Asjö, B., Goudsmit, J., Cupsa, A., Pascu, R., et al. (2000). Genetic evidence of multiple transmissions of HIV type 1 subtype F within Romania from adult blood donors to children. *AIDS Res. Hum. Retroviruses* 16, 327–336. doi: 10.1089/088922200309205

Palma, A. C., Araújo, F., Duque, V., Borges, F., Paixão, M. T., Camacho, R., et al. (2007). Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.* 7, 391–398. doi: 10.1016/j.meegid.2007.01.009

Paraschiv, S., Otelea, D., Baicus, C., Tinischi, M., Costache, M., and Neaga, E. (2009). Nucleoside reverse transcriptase inhibitor resistance mutations in subtype F1 strains isolated from heavily treated adolescents in Romania. *Int. J. Infect. Dis.* 13, 81–89. doi: 10.1016/j.ijid.2008.03.032

Paraskevis, D., Kostaki, E., Beloukas, A., Cañizares, A., Aguilera, A., Rodríguez, J., et al. (2015). Molecular characterization of HIV-1 infection in Northwest Spain (2009-2013): investigation of the subtype F outbreak. *Infect. Genet. Evol.* 16, 96–101. doi: 10.1016/j.meegid.2014.12.012

Paraskevis, D., Pybus, O., Magiorkinis, G., Hatzakis, A., Wensing, A. M., van de Vijver, D. A., et al. (2009). Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology* 6:49. doi: 10.1186/1742-4690-6-49

Peeters, M., D'Arc, M., and Delaporte, E. (2014). The origin and diversity of human retroviruses. *AIDS Rev.* 16, 23–34.

Pernas, B., Grandal, M., Mena, A., Castro-Iglesias, A., Cañizares, A., Wyles, D. L., et al. (2014). High prevalence of subtype F in newly diagnosed HIV-1 persons in northwest Spain and evidence for impaired treatment response. *AIDS* 28, 1837–1840. doi: 10.1097/QAD.0000000000000326

Pineda-Peña, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., et al. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* 19, 337–348. doi: 10.1016/j.meegid.2013.04.032

Pineda-Peña, A. C., Schrooten, Y., Vinken, L., Ferreira, F., Li, G., Trovão, N. S., et al. (2014). Trends and predictors of transmitted drug resistance (TDR) and clusters with TDR in a local Belgian HIV-1 epidemic. *PLoS One* 9:e101738. doi: 10.1371/journal.pone.0101738

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation in bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nat. Rev. Genet.* 5, 52–61. doi: 10.1038/nrg1246

Rodger, A. J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., van Lunzen, J., et al. (2016). Sexual activity without condoms and risk of HIV transmission in serodifferent couples when the HIV-positive partner is using suppressive antiretroviral therapy. *JAMA* 316, 171–181. doi: 10.1001/jama.2016.5148

Rodgers, M. A., Wilkinson, E., Vallari, A., McArthur, C., Sthreshley, L., Brennan, C. A., et al. (2017). Sensitive next-generation sequencing method reveals deep genetic diversity of HIV-1 in the Democratic Republic of the Congo. *J. Virol.* 91, e1841–e1816. doi: 10.1128/JVI.01841-16

Sayan, M., Willke, A., Ozgunes, N., and Sargin, F. (2013). HIV-1 subtypes and primary antiretroviral resistance mutations in antiretroviral therapy naïve HIV-1 infected individuals in Turkey. *Jpn. J. Infect. Dis.* 66, 306–311. doi: 10.7883/yoken.66.306

Schultz, A. K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., et al. (2009). jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* 37, W647–W651. doi: 10.1093/nar/gkp371

Snoeck, J., Van Laethem, K., Hermans, P., Van Wijngaerden, E., Derdelinckx, I., Schrooten, Y., et al. (2004). Rising prevalence of HIV-1 non-B subtypes in Belgium: 1983-2001. *J. Acquir. Immune Defic. Syndr.* 35, 279–285. doi: 10.1097/00126334-200403010-00009

Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. U.S.A.* 110, 228–233. doi: 10.1073/pnas.1207965110

Struck, D., Lawyer, G., Ternes, A. M., Schmit, J. C., and 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., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4:vey016. doi: 10.1093/ve/vey016

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Thomson, M. M., Fernández-García, A., Delgado, E., Vega, Y., Díez-Fuertes, F., Sánchez-Martínez, M., et al. (2012). Rapid expansion of a HIV-1 subtype F cluster of recent origin among men who have sex with men in Galicia, Spain. *J. Acquir. Immune Defic. Syndr.* 59, e49–e51. doi: 10.1097/QAI.0b013e3182400fc4

Triques, K., Bourgeois, A., Saragosti, S., Vidal, N., Mpoudi-Ngole, E., Nzilambi, N., et al. (1999). High diversity of HIV-1 subtype F strains in Central Africa. *Virology* 259, 99–109. doi: 10.1006/viro.1999.9720

Van Beckhoven, D., Florence, E., Ruelle, J., Deblonde, J., Verhofstede, C., Callens, S., et al. (2015). Good continuum of HIV care in Belgium despite weaknesses in retention and linkage to care among migrants. *BMC Infect. Dis.* 15:496. doi: 10.1186/s12879-015-1230-3

Vercauteren, J., Derdelinckx, I., Sasse, A., Bogaert, M., Ceunen, H., De Roo, A., et al. (2008). Prevalence and epidemiology of HIV type 1 drug resistance among newly diagnosed therapy-naive patients in Belgium from 2003 to 2006. *AIDS Res. Hum. Retroviruses* 24, 355–362. doi: 10.1089/aid.2007.0212

Verhofstede, C., Fransen, K., Van Den Heuvel, A., Van Laethem, K., Ruelle, J., Vancutsem, E., et al. (2017). Decision tree for accurate infection timing in individuals newly diagnosed with HIV-1 infection. *BMC Infect. Dis.* 17:738. doi: 10.1186/s12879-017-2850-6

Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., et al. (2000). Unprecedented degree of human immunodefiency 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. doi: 10.1128/JVI.74.22.10498-10507.2000

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., et al. (2016). 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539, 98–101. doi: 10.1038/nature19827

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 April 2019
doi: 10.3389/fmicb.2019.00655

Diverse Large HIV-1 Non-subtype B Clusters Are Spreading Among Men Who Have Sex With Men in Spain

Elena Delgado[1], Sonia Benito[1], Vanessa Montero[1], María Teresa Cuevas[1], Aurora Fernández-García[1,2], Mónica Sánchez-Martínez[1], Elena García-Bodas[1], Francisco Díez-Fuertes[3], Horacio Gil[1,4], Javier Cañada[1], Cristina Carrera[1], Jesús Martínez-López[1], Marcos Sintes[1], Lucía Pérez-Álvarez[1], Michael M. Thomson[1] and the Spanish Group for the Study of New HIV Diagnoses*

[1] *HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain,* [2] *CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain,* [3] *AIDS Immunopathogenesis Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain,* [4] *European Program for Public Health Microbiology Training, European Centre for Disease Prevention and Control, Stockholm, Sweden*

In Western Europe, the HIV-1 epidemic among men who have sex with men (MSM) is dominated by subtype B. However, recently, other genetic forms have been reported to circulate in this population, as evidenced by their grouping in clusters predominantly comprising European individuals. Here we describe four large HIV-1 non-subtype B clusters spreading among MSM in Spain. Samples were collected in 9 regions. A pol fragment was amplified from plasma RNA or blood-extracted DNA. Phylogenetic analyses were performed via maximum likelihood, including database sequences of the same genetic forms as the identified clusters. Times and locations of the most recent common ancestors (MRCA) of clusters were estimated with a Bayesian method. Five large non-subtype B clusters associated with MSM were identified. The largest one, of F1 subtype, was reported previously. The other four were of CRF02_AG (CRF02_1; n = 115) and subtypes A1 (A1_1; n = 66), F1 (F1_3; n = 36), and C (C_7; n = 17). Most individuals belonging to them had been diagnosed of HIV-1 infection in the last 10 years. Each cluster comprised viruses from 3 to 8 Spanish regions and also comprised or was related to viruses from other countries: CRF02_1 comprised a Japanese subcluster and viruses from 8 other countries from Western Europe, Asia, and South America; A1_1 comprised viruses from Portugal, United Kingom, and United States, and was related to the A1 strain circulating in Greece, Albania and Cyprus; F1_3 was related to viruses from Romania; and C_7 comprised viruses from Portugal and was related to a virus from Mozambique. A subcluster within CRF02_1 was associated with heterosexual transmission. Near full-length genomes of each cluster were of uniform genetic form. Times of MRCAs of CRF02_1, A1_1, F1_3, and C_7 were estimated around 1986, 1989, 2013, and 1983, respectively. MRCA locations for CRF02_1 and A1_1 were uncertain (however initial expansions in Spain in Madrid and Vigo, respectively, were estimated) and were most probable in Bilbao, Spain, for F1_3 and Portugal for C_7. These results show that the HIV-1 epidemic among MSM in Spain is becoming increasingly diverse through the expansion of diverse non-subtype B clusters, comprising or related to viruses circulating in other countries.

Keywords: HIV-1, molecular epidemiology, phylogeny, phylodynamics, men who have sex with men, subtypes, circulating recombinant forms, clusters

OPEN ACCESS

Edited by:
Gkikas Magiorkinis, National and Kapodistrian University of Athens, Greece

Reviewed by:
Apostolos Beloukas, University of West Attica, Greece
Juan Angel Patino Galindo, Columbia University, United States

****Correspondence:***
Michael M. Thomson
mthomson@isciii.es

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *25 November 2018*
Accepted: *15 March 2019*
Published: *03 April 2019*

Citation:
Delgado E, Benito S, Montero V, Cuevas MT, Fernández-García A, Sánchez-Martínez M, García-Bodas E, Díez-Fuertes F, Gil H, Cañada J, Carrera C, Martínez-López J, Sintes M, Pérez-Álvarez L, Thomson MM and the Spanish Group for the Study of New HIV Diagnoses (2019) Diverse Large HIV-1 Non-subtype B Clusters Are Spreading Among Men Who Have Sex With Men in Spain. Front. Microbiol. 10:655. doi: 10.3389/fmicb.2019.00655

INTRODUCTION

HIV-1 exhibits a characteristic high genetic variability, derived from elevated mutation and recombination rates. Through these mechanisms, the main (M) HIV-1 group, causative of the pandemic, has evolved into multiple genetic forms, designated subtypes, of which nine have been identified, subsubtypes, circulating recombinant forms (CRFs), of which 93 are currently recognized (HIV Sequence Database, 2019; Reis et al., 2019), and unique recombinant forms (URFs). The most globally prevalent HIV-1 genetic form is subtype C, estimated to represent around 47% worldwide infections, followed, in this order, by subtype B, subtype A, CRF02_AG, CRF01_AE, subtype G, and subtype D, with each of the remaining genetic forms estimated to represent less than 1% of global infections (Hemelaar et al., 2018).

In Western Europe, the predominant HIV-1 genetic form is subtype B, which was initially introduced among MSM and persons who inject drugs (PWID) (Lukashov et al., 1996; Casado et al., 2000; Kuiken et al., 2000; Hué et al., 2005; Beloukas et al., 2016). In early descriptions of HIV-1 genetic diversity in Western Europe, non-subtype B genetic forms were restricted to heterosexually infected immigrants coming from areas where those clades predominate, mainly sub-Saharan Africans, and European individuals epidemiologically linked to people from such areas (Fransen et al., 1996; Op de Coul et al., 1998; Thomson and Nájera, 2001). The first reports of HIV-1 non-subtype B genetic forms circulating in Western Europe among individuals without known epidemiological links to other geographic areas described the circulation of CRF01_AE among PWID in Finland (Liitsola et al., 2000) and of subtype G and CRF14_BG among a minority of HIV-1-infected PWID in the region of Galicia, Northwest Spain (Thomson et al., 2001; Delgado et al., 2002). Subsequent studies showed that the genetic forms circulating in Galicia derived from a subtype G variant widely circulating in Portugal, transmitted both through sexual contact and among PWID (Esteves et al., 2002, 2003; Palma et al., 2007; Carvalho et al., 2015). In recent years, an increasing prevalence of non-subtype B infections has been observed in Western Europe, reflecting both their importation from other geographical areas and their circulation among the local population (Abecasis et al., 2013; Beloukas et al., 2016; Hemelaar et al., 2018; Paraskevis et al., 2019).

In the current HIV-1 epidemic in Western Europe, the predominant propagation mode is through sexual contact among MSM (European Centre for Disease Prevention and Control, 2017 and WHO Regional Office for Europe 2017; Núñez et al., 2018), a population in which a resurgence of the HIV-1 epidemic has been observed since the 2000s, which is part of a global phenomenon (Bezemer et al., 2008; Beyrer et al., 2012). This has been accompanied by the emergence of phylogenetically identifiable transmission clusters, whose expansion is mostly driven by individuals with recent infection who are unaware of their HIV status (Lewis et al., 2008; Cuevas et al., 2009; Bezemer et al., 2010; Chalmet et al., 2010; Fisher et al., 2010; Zehender et al., 2010; Ambrosioni et al., 2012; Frange et al., 2012; Thomson et al., 2012; Audelin et al., 2013; Delgado et al., 2015; Esbjörnsson et al., 2016; Hoenigl et al., 2016; Chaillon et al., 2017; Patiño-Galindo et al., 2017; Parczewski et al., 2017; Verhofstede et al., 2018; Paraskevis et al., 2019). As expected, most clusters are of subtype B, but multiple instances of propagation of other HIV-1 genetic forms among European MSM have also been reported. These include subtypes A1 (Lai et al., 2016; Ragonnet-Cronin et al., 2016), C (de Oliveira et al., 2010; Lai et al., 2014; Ragonnet-Cronin et al., 2016), and F1 (Castro et al., 2010; Lai et al., 2012; Thomson et al., 2012; Delgado et al., 2015; Vinken et al., 2017; Verhofstede et al., 2018); CRF01_AE (von Wyl et al., 2011), CRF02_AG (Giuliani et al., 2013; Brand et al., 2014; Dauwe et al., 2015; Tamalet et al., 2015; Chaillon et al., 2017; Verhofstede et al., 2018), CRF17_BF (Fabeni et al., 2015), CRF19_cpx (Patiño-Galindo et al., 2015; González-Domenech et al., 2018; Pérez-Parra et al., 2018), CRF50_A1D (Foster et al., 2014), CRF56_cpx (Leoz et al., 2013), and CRF60_BC (Monno et al., 2012; Simonetti et al., 2014). However, the expansion of these clades has had a limited impact on the overall genetic composition of the HIV-1 epidemic among MSM in Western Europe, which is still largely dominated by subtype B. The only exception, though geographically restricted, is a large F1 subtype cluster of Brazilian ancestry, which represented a substantial proportion of new HIV-1 diagnoses among MSM in Northwest Spain (Thomson et al., 2012; Delgado et al., 2015). Here we describe four additional large non-subtype B clusters expanding among MSM in Spain, of CRF02_AG and of subtypes A1, F1 and C, each circulating in several Spanish regions and related to viruses from other countries.

Abbreviations: MCC, maximum clade credibility; MCMC, Markov chain Monte Carlo; ML, maximum likelihood; MRCA, most recent common ancestor; MSM, men who have sex with men; NFLG, near full-length genome; PP, posterior probability; PR-RT, protease-reverse transcriptase; PWID, people who inject drugs; tMRCA, time of most recent common ancestor.

MATERIALS AND METHODS

Samples

Plasma or whole blood samples were collected from 1999 to 2018 from HIV-1-infected individuals attended at hospitals from 15 provinces from 9 regions of Spain (Basque Country, Galicia, Navarre, Castilla y León, La Rioja, Madrid, Castilla-La Mancha, Aragón, and Comunidad Valenciana). The regional sample representativeness is variable, being the greatest in the regions of Basque Country, where all public hospitals participated, and Galicia, where all but one public hospitals participated. The study was approved by the Bioethics and Animal Well-being Committee of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain. Written informed consent was obtained from all participants in the study.

RNA and DNA Extraction, RT-PCR Amplification, and Sequencing

Amplification of HIV-1 fragments was done either from plasma RNA or from DNA extracted from whole blood. RNA was extracted from 1 ml plasma using Nuclisens EasyMAG kit (bioMérieux, Marcy l'Etoile, France), following the manufacturer's instructions. DNA was extracted from 200 μl

blood using QIAmp DNA Blood mini kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. An HIV-1 PR-RT fragment (approximately 1.4 kb) was amplified by RT-PCR followed by nested PCR, in the case of RNA, or by nested PCR, in the case of DNA, using previously reported primers (Delgado et al., 2015). In selected samples, NFLG amplification was done in four overlapping segments, as described (Delgado et al., 2002; Sierra et al., 2005), using RNA extracted either from plasma or from the primary isolate's culture supernatant grown from plasma using a previously described protocol (Delgado et al., 2012). Direct sequencing of the amplified products was done using an automated capillary sequencer. Sequence electropherograms were assembled and edited with Seqman (DNASTAR, Madison, WI, United States). Newly obtained sequences are deposited in GenBank under accessions MK177651-MK177824 (PR-RT sequences) and MK177825-MK177829, KT276258, KY496622, and KY989952 (NFLG sequences).

Phylogenetic Sequence Analyses

Sequences were aligned with MAFFT v.7 (Katoh and Standley, 2013). Initial trees with all sequences were constructed with the approximate ML method implemented in FastTree v2.1.10 (Price et al., 2010) using the general time reversible (GTR)+CAT evolutionary model and assessment of node support with Shimodaira-Hasegawa (SH)-like local support values. Transmission clusters were defined as those supported by SH-like values $\geq$ 0.95 comprising four or more individuals, with a majority of them being native Spanish. To determine whether the identified clusters were still supported when globally sampled sequences were included and to identify viruses from other areas belonging or phylogenetically related to them, the analyses were repeated including all PR-RT sequences of 1 kb or longer of the same genetic form available at the Los Alamos HIV Sequence Database (HIV Sequence Database, 2019), downloaded with the option "one sequence per patient." Support for clusters thus identified with FastTree was subsequently assessed with phylogenetic trees constructed via ML with PhyML v3.0 (Guindon et al., 2005), using the best-fit evolutionary model selected by Smart Model Selection (SMS) program (Lefort et al., 2017) and heuristic searches based on subtree pruning and regrafting (SPR) moves, with estimation of node support by the approximate likelihood ratio test (aLRT), SH-like procedure (Anisimova and Gascuel, 2006; Guindon et al., 2010). To keep computational times reasonable, in the analyses with PhyML only several hundred sequences (around 200–400) branching most closely to the clusters of interest in the previous FastTree analyses, together with PR-RT sequences from NFLG from databases, were included. Clusters were confirmed if in the analyses including database sequences with both FastTree and PhyML their node support values were $\geq$ 0.95. Trees were visualized with Dendroscope 3 (Huson and Scornavacca, 2012) or FigTree v1.4.2[1]. Intersubtype recombination was analyzed by bootscanning with Simplot v3.5 (Lole et al., 1999), with tree construction with the neighbor-joining method, using the Kimura 2-parameter substitution model and windows of 400 or 600 nt moving in 20 nt steps.

Antiretroviral Drug Resistance Determination

Antiretroviral (ARV) drug resistance was analyzed with the Calibrated Population Resistance Tool (Gifford et al., 2009).

Temporal and Geographic Estimations of Cluster Origins

To estimate times of the most recent common ancestors (tMRCA) of clusters and their most probable geographical locations, a Bayesian MCMC coalescent method as implemented in BEAST v1.8.1 (Drummond et al., 2012) was used. Prior to these analyses, the existence of temporal signal in the datasets was assessed by an analysis of root-to-tip distances against dates of sampling using TempEst (Rambaut et al., 2016). For the BEAST analyses, we used all PR-RT sequences $\geq$ 1 kb from each cluster and related sequences, as determined in the ML phylogenetic analyses, excluding sequences without data on year or location of sample collection. PR-RT sequences derived from NFLG sequences of the corresponding genetic form downloaded from the HIV Sequence Database were also included in these analyses, using not more than 5 sequences per country of sampling. In the case of the CRF02_1 cluster, 40 randomly selected CRF02_AG PR-RT sequences lacking drug resistance mutations downloaded from the Los Alamos HIV Sequences database were included, since using those derived from NFLG resulted in relatively low r^2 values in the TempEst analysis. We chose an HKY substitution model with gamma-distributed among-site rate heterogeneity and two partitions in codon positions (1st+2nd; 3rd) (Shapiro et al., 2006); uniform priors were used for absolute substitution rates (0-0.02 sub/site/year) and relative substitution rates in codon positions 1st+2nd and 3rd (0-10); we also used an uncorrelated lognormal relaxed clock model and a Bayesian skyline plot demographic model (Drummond et al., 2005). Each MCMC chain was run for 100 million to 200 million generations, sampling every 5,000 generations. MCMC convergence and effective sample sizes (ESS) were checked with Tracer v.1.6[2], ensuring that the ESS of each parameter was > 200. Results were summarized with a maximum clade credibility (MCC) tree, using TreeAnnotator v1.8.1, after removal of a 50% burn-in. The MCC trees were visualized with FigTree v1.4.2.[1] Parameter uncertainty was summarized in the 95% highest posterior density (HPD) intervals.

Statistical Analyses

Differences in clustering frequency between MSM and heterosexually infected individuals and changes in proportions of non-subtype B infections along time in newly diagnosed sexually infected individuals were analyzed with Fisher's exact test. Only native Spanish individuals were included in these analyses in order to focus on locally circulating strains, minimizing the

[1]http://tree.bio.ed.ac.uk/software/figtree/

[2]http://tree.bio.ed.ac.uk/software/tracer/

confounding effect of imported HIV-1 variants acquired in other countries. Numbers of individuals used in these analyses were 7080 in the first and 2060 in the second.

RESULTS

Using HIV-1 PR-RT sequences from 10,506 individuals obtained by us (whose data are summarized in **Supplementary Table 1**), we identified 320 phylogenetic clusters comprising 4 or more individuals, 247 of subtype B and 73 of other genetic forms. Belonging to a cluster was more frequent among Spanish MSM than among heterosexually-infected Spanish individuals (68% vs. 31%; $p = 7.7 \times 10^{-6}$) (**Supplementary Figure 1**). Differences were also significant when subtype B and non-subtype B clusters were analyzed separately. An increase along time of non-subtype B infections among newly diagnosed Spanish individuals was also observed. Statistically significant increases were observed in Spanish MSM from 2005–2009 to 2010–2014 (from 10.9% to 26.7%; $p = 1.4 \times 10^{-6}$) and among heterosexually infected Spanish individuals from 2010–2014 to 2015–2018 (from 28.5% to 39.7%; $p = 0.0097$) (**Supplementary Figure 2**).

Of the non-subtype B clusters, 5 large ones (here defined as those comprising 10 or more individuals) were associated with MSM. The largest one, of F1 subtype (currently comprising 187 individuals), was reported previously (Thomson et al., 2012; Delgado et al., 2015). The other four were of CRF02_AG and of subtypes A1, F1, and C, henceforth designated CRF02_1, A1_1, F1_3, and C_7, respectively. All four clusters were well supported when the analyses were repeated with FastTree including all PR-RT sequences > 1 kb of the respective genetic forms available at the Los Alamos HIV Sequence database (with numbers from 1,318 of subtype F1 to 22,762 of subtype C) and with PhyML including several hundred (from 312 for subtype C to 513 for CRF02_AG) database sequences branching closer to each cluster in the FastTree analysis and derived from NFLG sequences. These analyses also allowed to identify viruses from databases belonging or closely related to the clusters. Epidemiological data of samples studied by us belonging to these clusters and to subclusters within them are shown in **Table 1**.

CRF02_1 Cluster

CRF02_1 comprised 115 individuals, 67 studied by us and 48 whose sequences were retrieved from databases (**Figure 1**). Most samples were collected in Spain, but there were also samples from Japan, Switzerland, United Kingdom, Ecuador, Netherlands, Sweden, Germany, Malaysia, and Hong Kong. Samples from Spain were from 8 regions, mainly from Basque Country, Madrid, and Aragon.

Years of HIV-1 diagnosis were 2008–2018.

CRF02_1 comprised three subclusters, designated with numerals 1-3. CRF02_1_1 comprised 60 individuals from 7 Spanish regions, two from United Kingdom and one from Germany. Within it, two subsubclusters were distinguished, associated with the Basque Country and the city of Zaragoza, Aragón, respectively; CRF02_1_2 comprised 13 individuals from Japan; and CRF02_1_3 comprised 11 individuals, 10 of them from Zaragoza.

Among samples studied by us, 88% were from men. Transmission route was sexual in the great majority, and, among those infected via sexual contact, 51% were self-declared MSM. Interestingly, all 10 individuals from subcluster CRF02_1_3 studied by us (5 women and 5 men, all from Zaragoza) were infected via heterosexual contact.

Most individuals in this and in the other clusters here described were native Spanish (**Table 1**).

TABLE 1 | Epidemiological data of clusters*.

Cluster/ Subcluster	N	Gender			Transmission route					Region									Country of origin†				Years of HIV-1 diagnosis
		M	F	ND	MSM	MSx	HT	PWID	ND	BC	GA	NA	MD	AR	LR	CV	CL	CM	Spain	LAm	Other	ND	
CRF02_1	67	58	8	1	30	12	18	1	6	31	3	3	10	15		3	1	1	53	10	4		2008–2018
1	49	47	1	1	28	12	5	1	3	26	3	1	9	6		3		1	40	7	2		2008–2018
1_1	17	17			11	4	1		1	15	2								14	2	1		2008–2018
1_2	7	7			3	1	3						1	6					6	1			2016–2018
3	10	5	5				10							9			1		8	1	1		2016–2018
A1_1	54	50	3	1	36	6	6		6	6	22	18	3		5				37	10	5	2	2006–2018
1	30	27	3		18	4	3		5	6		18	1		5				19	5	5	1	2013–2017
2	16	15		1	12	1	2	1			15		1						13	2		1	2009–2018
2_1	10	10			8		1		1		9		1										2016–2018
F1_3	36	36			17	10	3		6	30		3	3						29	6	1		2014–2018
C_7	14	14			5	6	3			4	9						1		10	1		3	2009–2017
1	8	8			5	2	1				7								6	1		1	2012–2017
2	4	4				3	1			4									3			1	2009–2015

*Only data from samples processed by the authors are included. †Country of which the individual is native. ND, no data. HT, heterosexual. MSx, male, unspecified sexual. BC, Basque Country. MD, Madrid. GA, Galicia. NA, Navarre. AR, Aragon. CL, Castilla y León. CM, Castilla-La Mancha. CV, Comunidad Valenciana. LR, La Rioja. LAm, Latin American country.

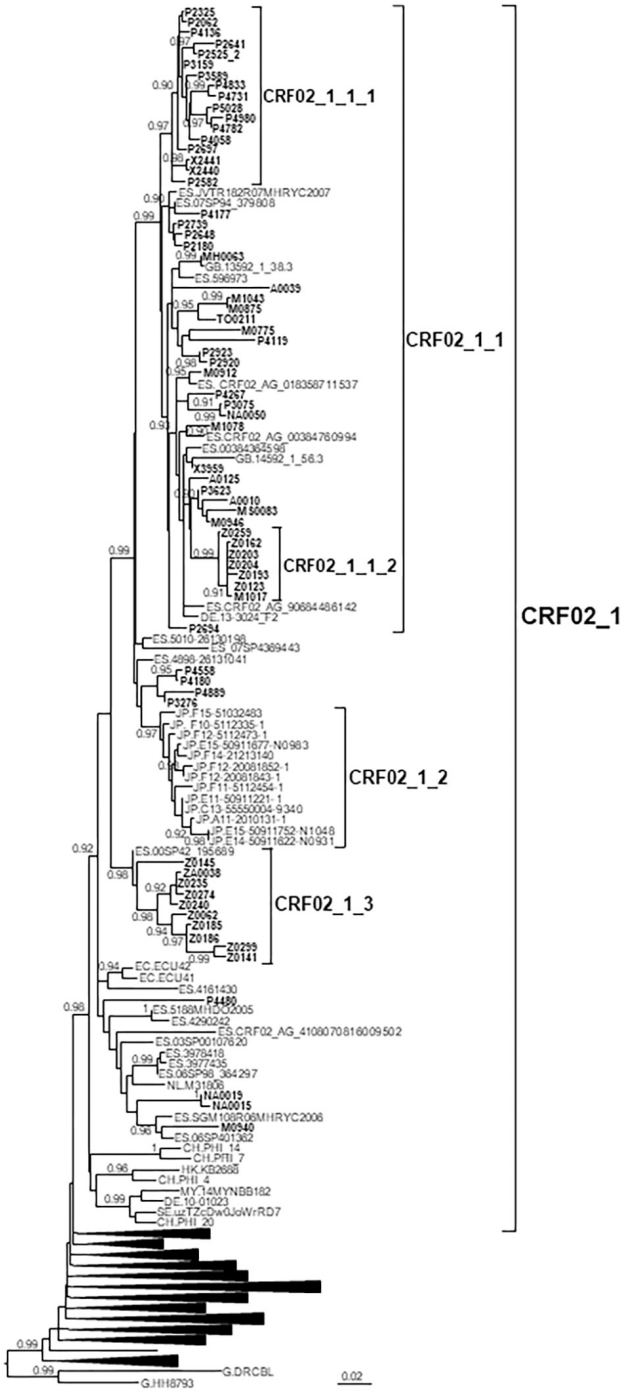


FIGURE 1 | Maximum likelihood tree of PR-RT sequences of the CRF02_1 cluster. The tree was constructed with PhyML, with assessment of node support with the aLRT SH-like procedure. The analysis incorporates 513 CRF02_AG PR-RT sequences from databases that in preliminary analyses with FastTree branched closer to the CRF02_1 cluster and from NFLG sequences, and two subtype G sequences used to root the tree. For better viewing, clades outside of the CRF02_1 cluster are collapsed. Only aLRT SH-like node support values ≥ 0.9 are shown. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of the country of sample collection followed by the sample name.

A1_1 Cluster

A1_1 comprised 66 individuals, 54 studied by us and 12 with sequences deposited in databases (7 from United Kingdom, 4 from Portugal, and 1 from United States) (**Figure 2**). A1_1 was also related to viruses from the A1 subtype lineage circulating in Greece, Albania, and Cyprus (Ciccozzi et al., 2005; Paraskevis et al., 2007; Pineda-Peña et al., 2018). Samples from Spain were collected in 5 regions, mainly in Galicia and Navarre.

Forty five of 47 individuals with available data were diagnosed in 2012–2018.

Within A1_1, there were two main subclusters: A1_1_1, comprising all individuals from Navarre, Basque Country, and La Rioja, and 1 from Madrid; and, A1_1_2 comprising individuals mostly from Galicia, with a majority grouping in a subsubcluster.

Most (93%) individuals in the cluster were men, with predominance of MSM.

F1_3 Cluster

F1_3 comprised 36 individuals. There were no sequences from databases belonging to F1_3, but viruses from Romania, most of them transmitted sexually (Niculescu et al., 2015), were closely related to it (**Figure 3**). Five viruses from Spain (4 of them sequenced by us), 3 from United Kingdom, and 1 from Poland also branched close to F1_3, interspersed among the Romanian samples. Most F1_3 samples were from Basque Country.

All infections were diagnosed in 2014 or later. All individuals in F1_3 were men. Transmission route was sexual in all for which data were available, with a majority of MSM.

C_7 Cluster

C_7 comprised 17 individuals, including 14 studied by us and 3 from Portugal, whose sequences were retrieved from databases. One sequence from Mozambique branched basally to C_7 (**Figure 4**). C_7 comprised two subclusters, C_7_1, comprising all but one samples from Galicia, and C_7_2, comprising all 4 samples from Basque Country. All but one had been diagnosed in 2012–2017.

All but one Spanish samples were from Galicia or Basque Country. All individuals in C_7 were sexually infected men, with 5 being self-declared MSM.

Bayesian Analyses and Temporal and Geographic Estimations

To estimate the temporal and geographic origins of clusters and subclusters, Bayesian coalescent analyses were performed with PR-RT sequences, summarizing the posterior distribution of trees with MCC trees. Prior to these analyses, temporal signal was analyzed, revealing a clock-like structure in all four datasets used for subsequent analyses (r^2 = 0.4395 in CRF02_1, r^2 = 0.5399 in A1_1, r^2 = 0.7175 in F1_3, and r^2 = 0.3518 in C_7), indicating that the datasets contained sufficient temporal structure for reliable estimation of divergence times.

In the Bayesian analyses, all four clusters previously defined via ML were supported by node PP values > 0.98.

The tMRCA of the entire CRF02_1 cluster was estimated around 1986, but the location of the MRCA was uncertain,

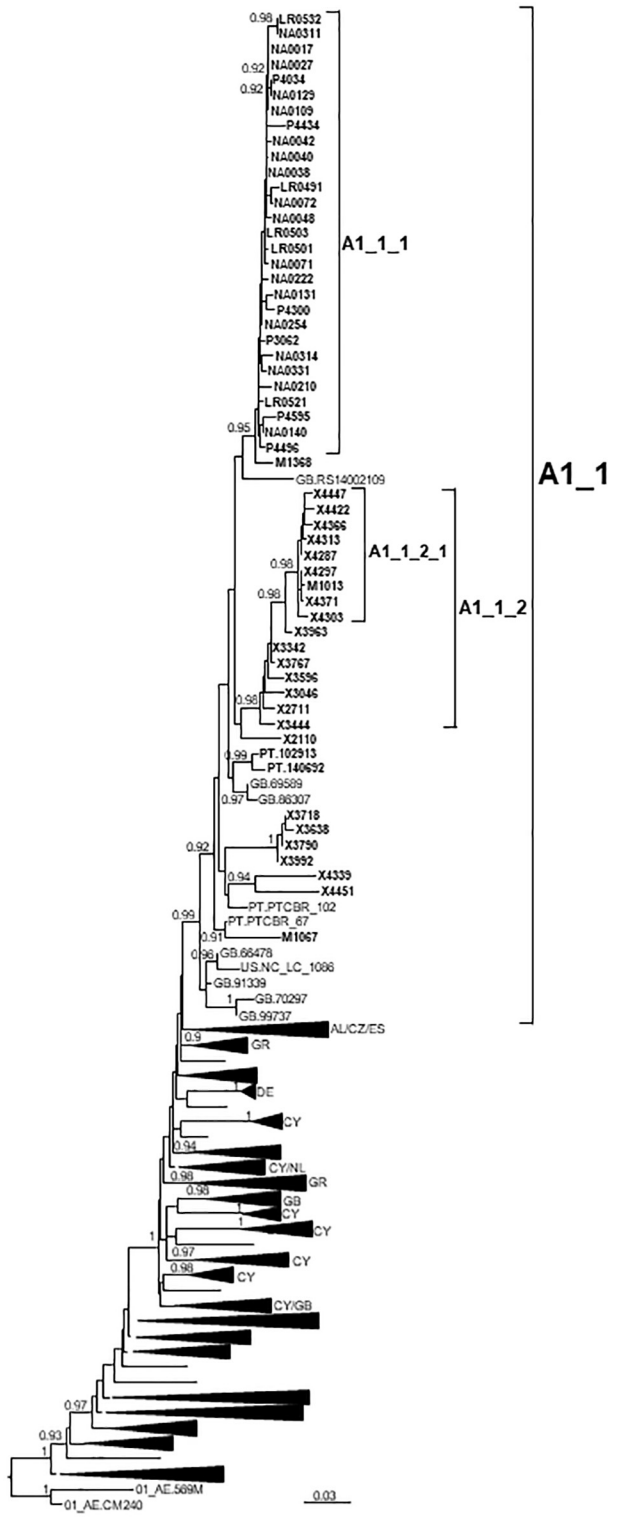


FIGURE 2 | Maximum likelihood tree of PR-RT sequences of the A1_1 cluster. The tree was constructed with PhyML, with assessment of node support with the aLRT SH-like procedure. The analysis incorporates 338 A1 subsubtype PR-RT sequences from databases that in preliminary analyses with FastTree branched closer to the A1_1 cluster and from NFLG sequences, and two *(Continued)*

FIGURE 2 | Continued
CRF01_AE sequences used to root the tree. For better viewing, clades outside of the A1_1 cluster are collapsed. Only aLRT SH-like node support values ≥ 0.90 are shown. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of the country of sample collection followed by the sample name. Collapsed clades most closely related to the A1_1 cluster are labeled with the two-letter codes of the countries of sample collection, excluding countries represented by a single sequence.

since the most probable one was supported by a PP < 0.5. However, the location of the MRCA of the subcluster comprising all but the 8 most basal sequences (collected in Switzerland, Sweden, Germany, Malaysia, and Hong Kong), with tMRCA around 1988, had a strong support in Madrid, with a PP of 0.975 (**Figure 5**). Subcluster CRF02_1_1 emerged around 2002 in Madrid, CRF02_1_2 around 2006 in Japan, and CRF02_1_3 around 2006 in Zaragoza.

The estimated tMRCA of the entire A1_1 cluster was around 1989, but the location of the MRCA was uncertain, since the most probable one was supported by a PP < 0.5. However, the location of the MRCA of the subcluster comprising all but the 6 most basal sequences, with tMRCA around 1994, was supported by a PP of 0.814 in the city of Vigo, Galicia. The emergence of subcluster A1_1_1 was around 2010, with highest location PP in Pamplona, and that of A1_1_2 was around 2004 in Vigo, Galicia (**Figure 6**). It should be pointed out that 2 samples from Portugal and 6 from United Kingdom could not be used in the Bayesian analyses, since no collection year was available for them.

The estimated tMRCA of F1_3 was around 2013 in the city of Bilbao, with a strongly supported ancestry in Romania (**Figure 7**).

Finally, the estimated tMRCA of C_7 was around 1983. Its most probable origin was Portugal, but with a PP support for the entire cluster of only 0.6. However, the location of the MRCA of the subcluster comprising all but the most basal sample had a strong support in Portugal (PP = 0.91) (**Figure 8**). Subcluster C_7_1 emerged around 2008 in Vigo and C_7_2 around 2004 in Vitoria, Basque Country.

ARV Drug Resistance Mutations

In CRF02_1, 5 sequences (4 from Spain and one from Germany) had ARV drug resistance mutations. One was from a drug-experienced individual in therapeutic failure with multiple drug resistance mutations, and two, with K101E and K103N, respectively, mutations of resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI), were from drug-naïve individuals. The other two, with K103N and K101E mutations, respectively, were from database sequences without data on drug treatment. In A1_1, one database sequence from United Kingdom had Y188C and G190A NNRTI resistance mutations. In C_7, all but 2 sequences had L90M mutation associated with protease inhibitor drug resistance; all 13 Spanish sequences with this mutation were from drug-naïve individuals.

Near Full-Length Genome Sequences

To determine whether the viruses from the clusters were of uniform genetic form all along their genomes or were interclade

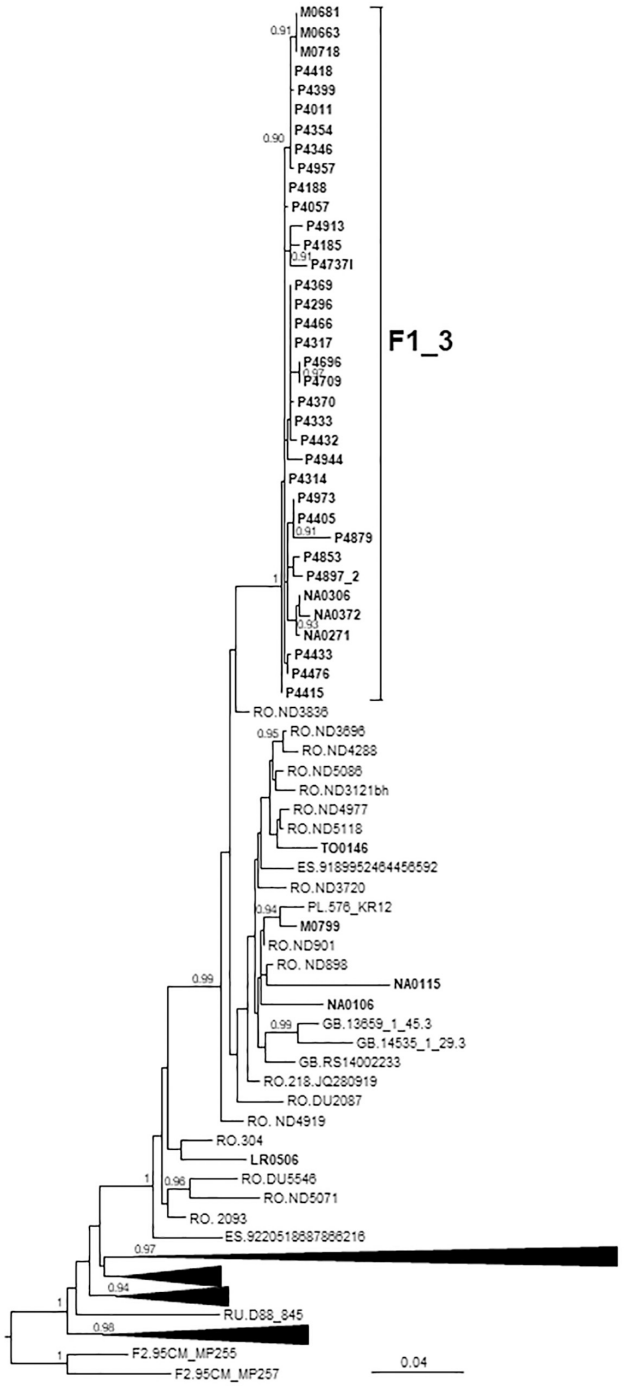


FIGURE 3 | Maximum likelihood tree of PR-RT of viruses of the F1_3 cluster. The tree was constructed with PhyML, with assessment of node support with the aLRT SH-like procedure. The analysis incorporates 358 PR-RT F1 subsubtype sequences from databases that in preliminary analyses with FastTree branched closer to the F1_3 cluster and from NFLG sequences, and two F2 subsubtype sequences used to root the tree. For better viewing, clades outside of the F1_3 cluster, excluding those most closely related to the F1_3 cluster, are collapsed. Only aLRT SH-like node support values $\geq$ 0.90 are shown. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of the country of sample collection followed by the sample name.

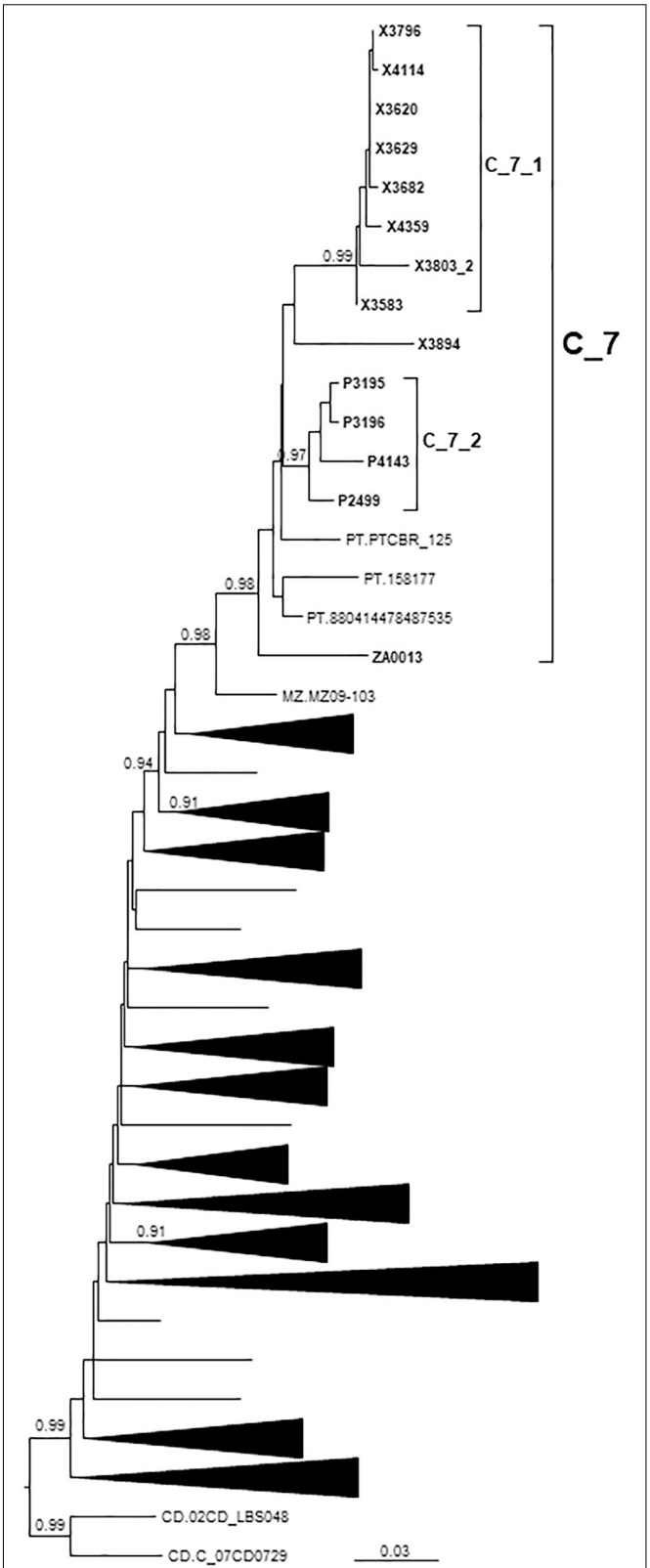


FIGURE 4 | Maximum likelihood tree of PR-RT of viruses of the C_7 cluster. The tree was constructed with PhyML, with assessment of node support with

(Continued)

FIGURE 4 | Continued
the aLRT SH-like procedure. The analysis incorporates 312 PR-RT subtype C sequences from databases that in preliminary analyses with FastTree branched closer to the C_7 cluster and from NFLG sequences, and two subtype C sequences from the Democratic Republic of Congo used to root the tree. For better viewing, clades outside of the C_7 cluster are collapsed. Only aLRT SH-like node support values ≥ 0.90 are shown. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of the country of sample collection followed by sample name.

recombinants, two NFLG sequences were obtained for each cluster, either from plasma RNA (P2648, P3075, P4496, P4346, and P4476) or from RNA extracted from culture supernatant (NA0048_2, X3303_2, and X3988). An additional NFLG from the A1_1 cluster (X2110, GenBank accession FJ670523) had been obtained previously by us (Cuevas et al., 2010). Bootscan analyses showed that all were of uniform genetic form along their genomes (**Figure 9**). We note that NFLG sequences of two CRF02_1 viruses from United Kingdom (**Figure 1**) are also available at sequence databases (Yebra et al., 2018; HIV Sequence Database, 2019) (GenBank accessions MF109381, MF109550).

DISCUSSION

In Western Europe, subtype B has been largely predominant among MSM since the early HIV-1 epidemic, but in recent years other genetic forms have been reported to be circulating in this population, as evidenced by their grouping in phylogenetic clusters comprising mostly European individuals. In this study, based on a large dataset from Spain, we found an increase in proportions of non-subtype B infections among MSM in recent years (**Supplementary Figure 2**) and higher clustering frequency in MSM compared to heterosexually infected individuals (**Supplementary Figure 1**). Among clusters associated with MSM, five large ones were of non-subtype B genetic forms, one of which, of F1 subtype, was reported previously (Thomson et al., 2012; Delgado et al., 2015). The other four were of CRF02_AG and subtypes A1, F1, and C, for which here we analyze epidemiological correlations, estimated emergence times and places, NFLGs, and drug resistance mutations.

The CRF02_AG cluster (CRF02_1) comprised 115 individuals, including 67 studied by us and 48 whose sequences were retrieved from databases, making it one of the largest non-subtype B clusters circulating among MSM reported to date in Western Europe (Delgado et al., 2015; Vinken et al., 2017). CRF02_AG is the predominant HIV-1 genetic form in most West African countries (Montavon et al., 2000; Hemelaar et al., 2018) and is common in West-Central Africa. It also propagates as a minor form in several Western European countries (Giuliani et al., 2013; Brand et al., 2014; Tamalet et al., 2015; Beloukas et al., 2016; Chaillon et al., 2017; Verhofstede et al., 2018), Tunisia (El Moussi et al., 2017), and Brazil (Delatorre et al., 2012), and in 2002 caused an outbreak among PWID in Uzbekistan (Carr et al., 2005), with subsequent dissemination to Kazakhstan (Eyzaguirre et al., 2007; Lapovok et al., 2014) and Russia (Moskaleychik et al., 2015), giving rise to CRF63_02A1 through recombination with the former Soviet Union subtype A variant (Baryshev et al., 2014; Shcherbakova et al., 2014). CRF02_AG has been reported to be one of the most common non-subtype B genetic forms in Western Europe (together with subtypes A1 and C) (Abecasis et al., 2013; Beloukas et al., 2016; Hemelaar et al., 2018) and in Spain (Yebra et al., 2012). The CRF02_AG cluster here described is not completely new, since a cluster of four individuals from the region of Valencia belonging to it was reported by other authors, who also noted that 9 sequences from databases, 7 from Spain and 2 from Ecuador, were related to the Valencian cluster (Bracho et al., 2014). However, the data here presented considerably enlarge the size and the geographic range of the cluster. CRF02_1 comprises viruses from 8 Spanish regions and from 9 other countries, from Western Europe, Asia and South America, with 13 Japanese viruses grouping in a monophyletic subcluster, indicating that it is circulating in this country. Although CRF02_1 is mainly associated with MSM, a subcluster comprising 11 individuals, 10 of them from the city of Zaragoza, propagates via heterosexual contact (**Table 1**). The origin of CRF02_1 is not recent, with a tMRCA estimated around 1986, with uncertain location, for the entire cluster, or 1988 in Madrid for the subcluster excluding the 8 most basal sequences, but its three major subclusters emerged in recent years, with tMRCAs in the 2000s.

The A1 subtype cluster (A1_1) is the second largest cluster here described, with 66 individuals, 54 studied by us and 12 with sequences in databases. A1 subtype circulates mainly in Eastern, Central and Western Africa (Hemelaar et al., 2018), all former Soviet Union (FSU) countries (Bobkova, 2013), Greece (Paraskevis et al., 2007), Albania (Ciccozzi et al., 2005), and Cyprus (Pineda-Peña et al., 2018), and as a minor form in India (Pandey et al., 2016), although some authors designate the variants circulating in Western Africa and FSU as distinct subsubtypes (A3 and A6, respectively) (Meloni et al., 2004; Foley et al., 2016). The lineage circulating in Greece and Albania, also detected in Cyprus, is of monophyletic origin, with estimated tMRCA around 1978 (Paraskevis et al., 2007). A1 subtype clusters have been reported in United Kingdom (Gifford et al., 2007; Hughes et al., 2009; Ragonnet-Cronin et al., 2016), Italy (Lai et al., 2016), Switzerland (von Wyl et al., 2011), and Portugal (Carvalho et al., 2015), but the one here reported is the largest reported to date in Western Europe. A1_1 comprises individuals from 5 Spanish regions and 3 other countries (United Kingdom, Portugal, and United States), and is related to the Greek-Albanian A1 lineage (**Figure 2**). Its origin is not recent, with estimated tMRCA around 1989, with uncertain location for the entire cluster, or around 1994 in Vigo, Galicia, excluding the 6 most basal sequences, but its two major subclusters are of recent origin, with tMRCAs around 2004 and 2010, respectively.

The F1 cluster (F1_3) comprises 36 individuals, all resident in Spain, most of them in Bilbao, Basque Country. Of the clusters here described, this is the one with the most recent origin, with estimated tMRCA around 2013 in Bilbao. It is currently increasing in size, with 6 individuals newly diagnosed in 2018. Subtype F1 is circulating in Central Africa, Brazil and Romania (Dumitrescu et al., 1994; Louwagie et al., 1994; Bandea et al., 1995;

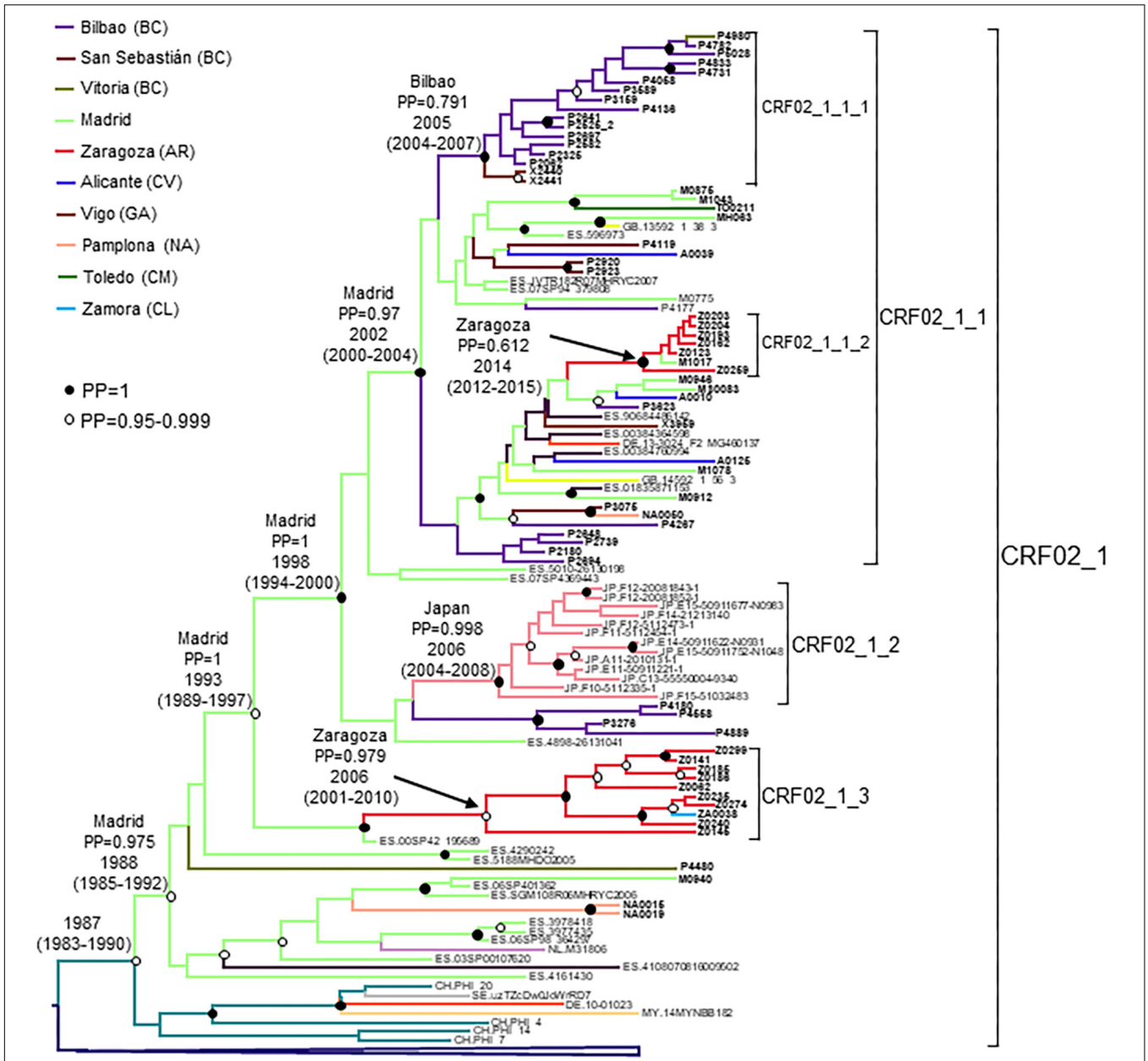


FIGURE 5 | Maximum clade credibility tree of PR-RT sequences of the CRF02_1 cluster. The tree also includes 40 other PR-RT CRF02_AG sequences from databases. For better viewing, the clade comprising viruses branching outside of the CRF02_1 cluster is collapsed. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of the country of sample collection followed by the sample name. Nodes supported by PP = 1 and PP = 0.95–0.999 are marked with filled and unfilled circles, respectively. Colors of terminal and internal branches represent sampling locations and most probable locations of the corresponding nodes, respectively, according to the legend on the left. For the nodes corresponding to CRF02_1 cluster and its major subclusters, the location with the highest PP (if > 0.5) and the tMRCAs (with 95% HPD intervals) are indicated above or close to the subtending branches.

Apetrei et al., 1997; Op de Coul et al., 2000; Hemelaar et al., 2018), and F1 subtype clusters have been recently identified in Spain (Thomson et al., 2012; Delgado et al., 2015), Belgium (Vinken et al., 2017; Verhofstede et al., 2018), Switzerland (Castro et al., 2010), Italy (Lai et al., 2012), and Portugal (Carvalho et al., 2015). F1_3 is most closely related to F1 viruses from Romania (**Figure 4**), which are related to viruses circulating in Angola (Guimarães et al., 2009) and initially propagated among adults via sexual contact, with subsequent propagation among institutionalized children through contaminated injection equipment (Op de Coul et al., 2000; Bello et al., 2012), and more recently among PWID (Temereanca et al., 2013; Niculescu et al., 2015). The expansion of an F1 subtype cluster of Romanian ancestry in Spain has its counterpart in the recent expansion of CRF14_BG, originally described in Spain (Thomson et al., 2001; Delgado et al., 2002) and Portugal (Esteves et al., 2003;

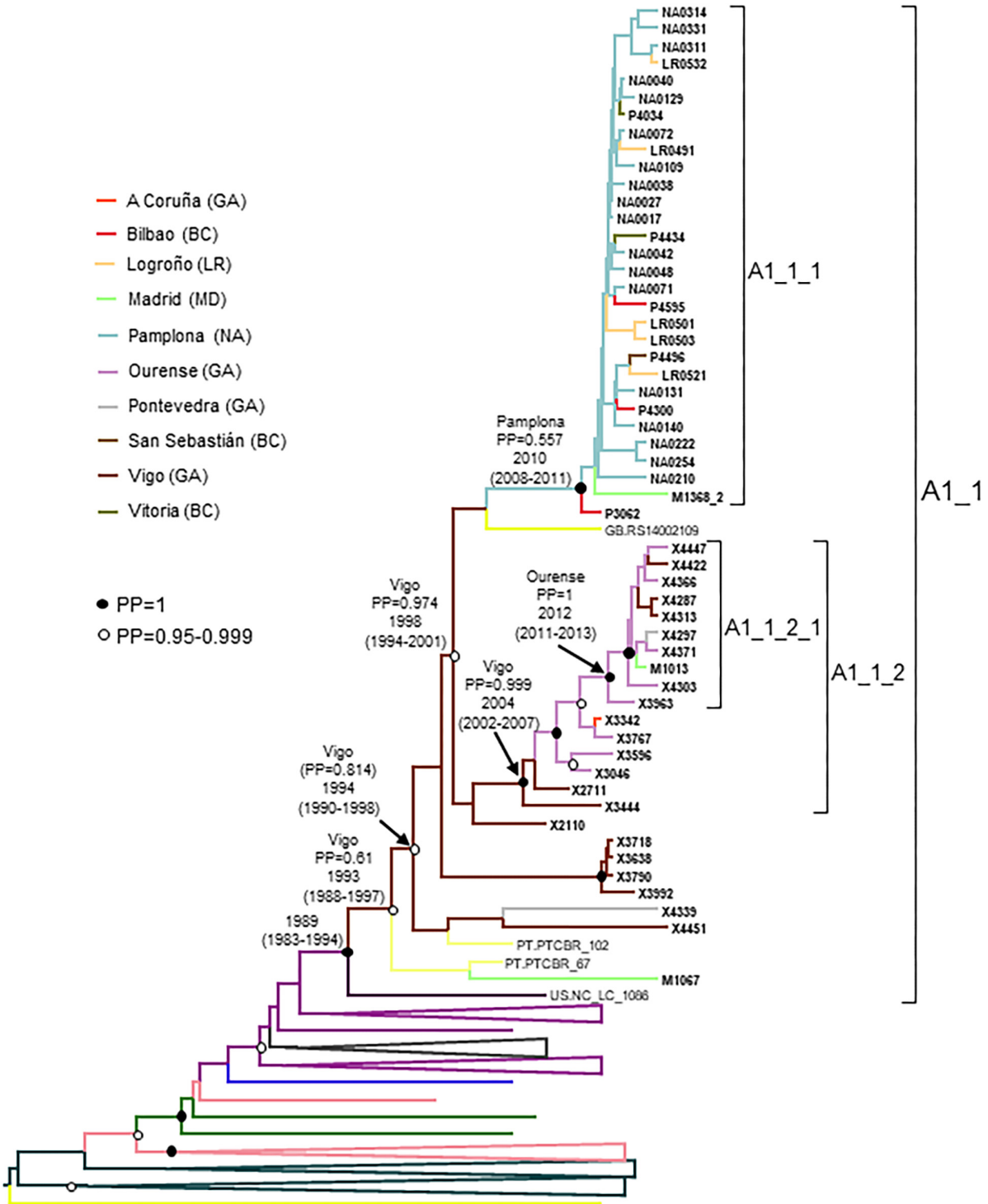


FIGURE 6 | Maximum clade credibility tree of PR-RT sequences of the A1_1 cluster. The tree also includes a sequence from US that in ML trees branched close to the A1_1 cluster (samples from United Kingdom were excluded because no information on time of sample collection was available), and A1 subsubtype PR-RT sequences from NFLG sequences from databases. For better viewing, clades outside of the A1_1 cluster are collapsed. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of country of sample collection followed by sample name. Clades most closely related to the A1_1 cluster are labeled with the two-letter ISO code of the sampling countries of viruses contained in it. Nodes supported by PP = 1 and PP = 0.95-0.999 are marked with filled and unfilled circles, respectively. Colors of terminal and internal branches represent sampling locations and most probable locations of the corresponding nodes, respectively, according to the legend on the left. For the nodes corresponding to the A1_1 cluster and its major subclusters, the location with the highest PP (if > 0.5) and the tMRCAs (with 95% HPD intervals) are indicated above or close to the subtending branches.

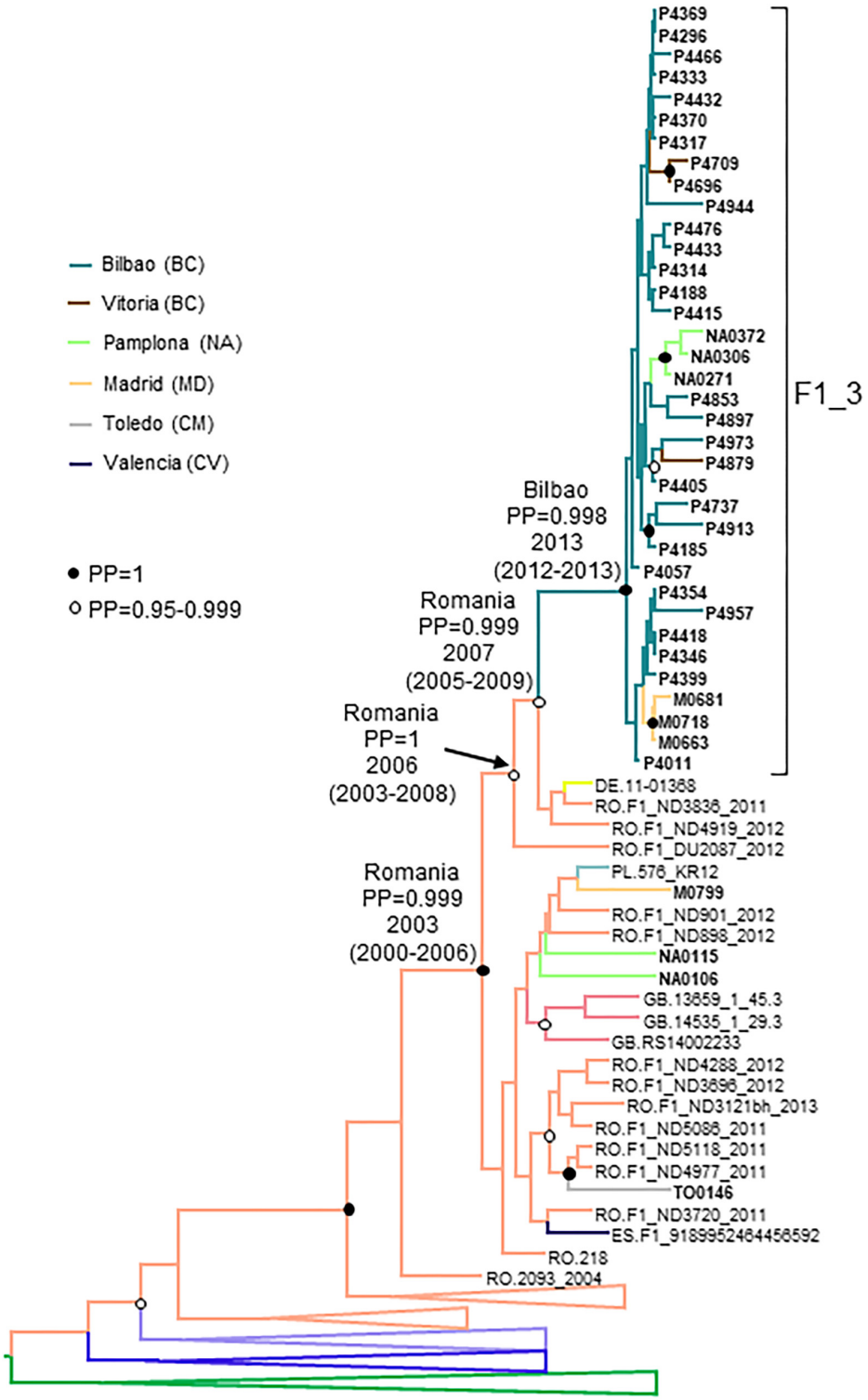


FIGURE 7 | Maximum clade credibility tree of PR-RT sequences of the F1_3 cluster. The tree also includes F1 subsubtype sequences from databases branching close to the F1_3 cluster and F1_3 subsubtype PR-RT sequences from NFLG sequences from databases. For better viewing, clades outside of the F1 cluster are collapsed, except those most closely related to the F1_3 cluster. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of country of sample collection followed by sample name. Nodes supported by PP = 1 and PP = 0.95–0.999 are marked with filled and unfilled circles, respectively. Colors of terminal and internal branches represent sampling locations and most probable locations of the corresponding nodes, respectively, according to the legend on the left. For the nodes corresponding to the F1_3 cluster and the clades within which it is contained, the location posterior probabilities and the tMRCAs (with 95% HPD intervals) are indicated above or close to the subtending branches.

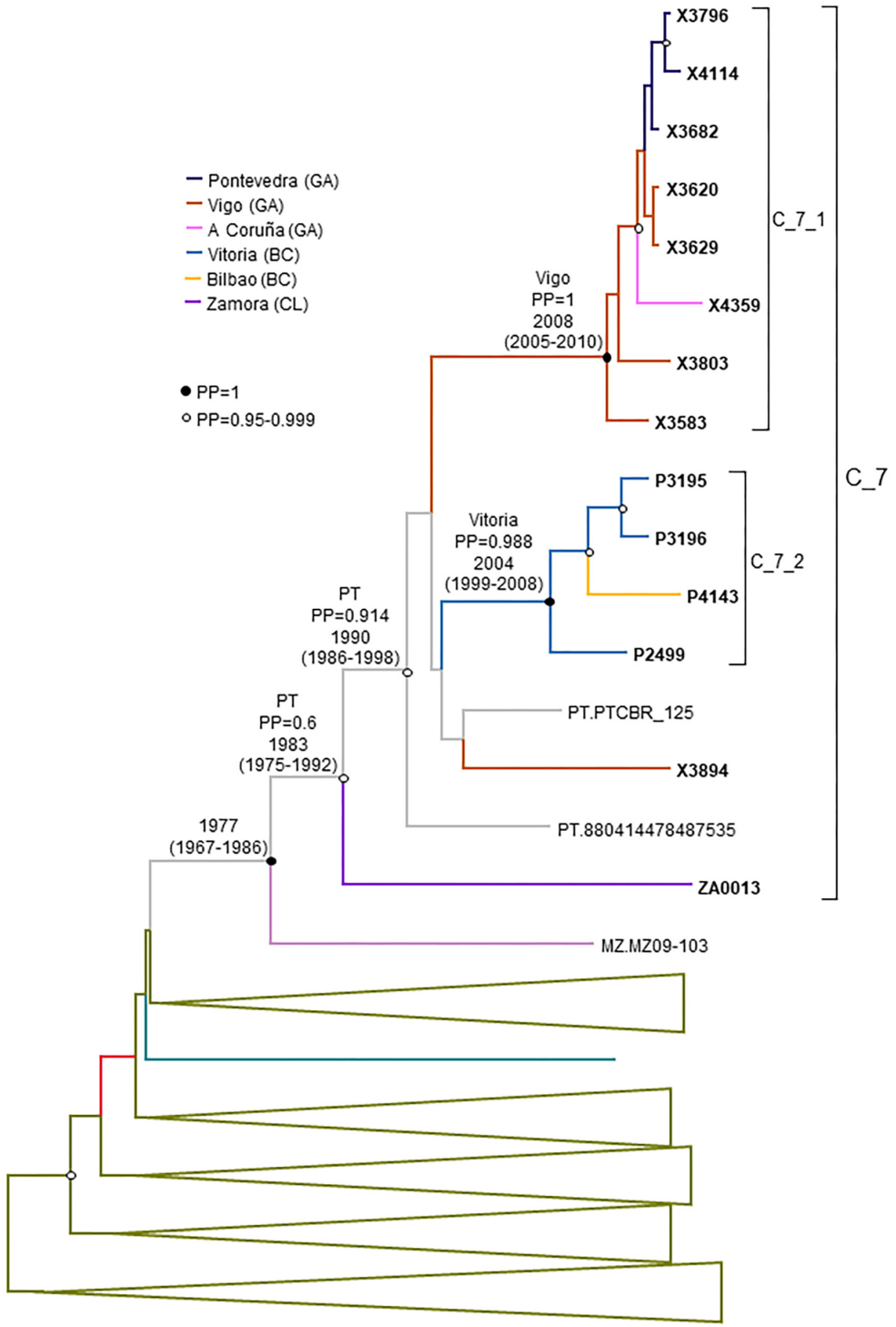


FIGURE 8 | Maximum clade credibility tree of PR-RT sequences of the C_7 cluster. The tree also includes a subtype C sequences from Mozambique that in ML trees branched close to the C_7 cluster and subtype C PR-RT sequences from NFLG sequences from databases. For better viewing, clades outside of the C_7 cluster are collapsed. Sequences obtained by us are in bold type. Sequences from databases are labeled with the two-letter ISO code of country of sample collection followed by sample name. Nodes supported by PP = 1 and PP = 0.95–0.999 are marked with filled and unfilled circles, respectively. Colors of terminal and internal branches represent sampling locations and most probable locations of the corresponding nodes, respectively, according to the legend on the left. For the nodes corresponding to the C_7 cluster and subclusters within it, the location posterior probability and the tMRCAs (with 95% HPD intervals) are indicated above or close to the subtending branches. tMRCA is also indicated for the node corresponding to the clade including the sample from Mozambique (most probable location is omitted, since its PP is below 0.5).

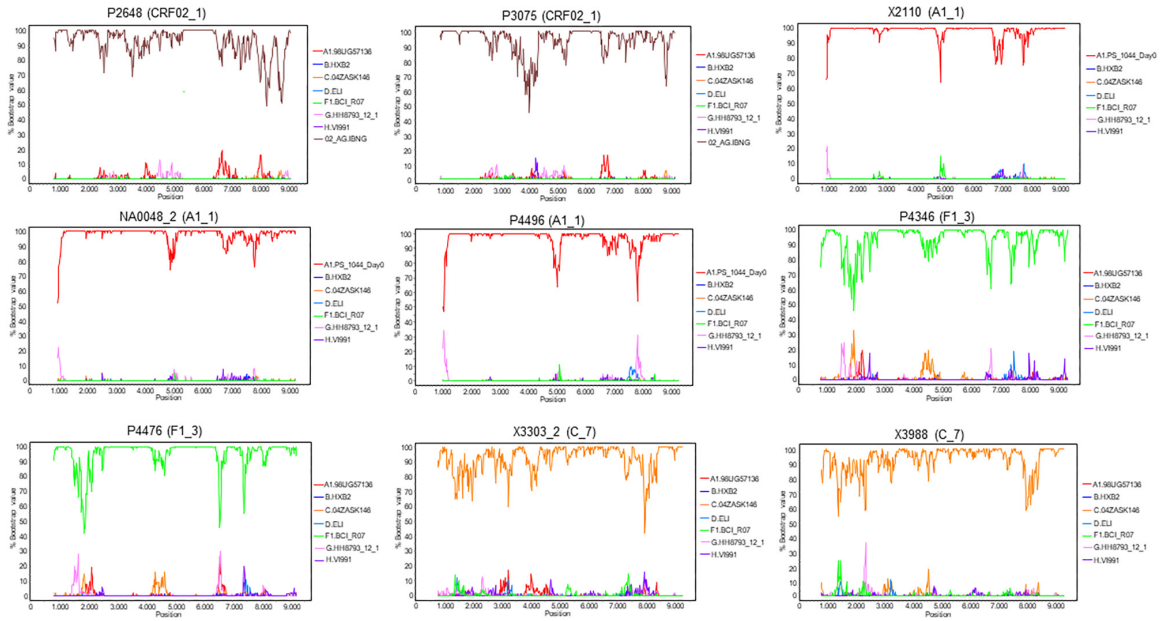


FIGURE 9 | Bootscan analyses of NFLG sequences of viruses from the identified clusters. Virus names, with cluster in parentheses, are above each plot. A window of 400 nt was used for viruses of the A1_1, F1_3, and C_7 clusters, and of 600 nt for viruses of the CRF02_1 cluster, sliding in 20 nt increments. The horizontal axis represents the nucleotide position in the HXB2 proviral genome of the window's midpoint. The vertical axis represents the bootstrap value supporting clustering with reference sequences, listed on the right of each plot.

Abecasis et al., 2011; Bartolo et al., 2011), among PWID in Romania (Niculescu et al., 2015). The exchange of HIV-1 genetic forms between Romania and Spain reflects the presence of a large Romanian immigrant population in Spain (Instituto Nacional de Estadística, 2018), frequently traveling between both countries.

The subtype C cluster (C_7) comprises 17 viruses, 14 from three Spanish regions, predominantly from Galicia, and three from Portugal. It comprises a Galician and a Basque subclusters (**Figure 4**). This is not the first report of this cluster, as it was described by us when it comprised only 7 individuals, within the context of the description of HIV-1 clusters bearing ARV drug resistance mutations, noting that viruses belonging to it carry the L90M mutation of resistance to protease inhibitors (Vega et al., 2015). This mutation was found in 15 of 17 C_7 viruses. Subtype C is the most prevalent clade in the HIV-1 pandemic, circulating mainly in Southern and Eastern Africa, Southern Brazil, and South Asia (Hemelaar et al., 2018). In Western Europe, subtype C clusters have been reported in the United Kingdom (Hughes et al., 2009; de Oliveira et al., 2010; Ragonnet-Cronin et al., 2016), Italy (Lai et al., 2014), and Portugal (Carvalho et al., 2015).

The origin of C_7 is not recent, with a tMRCA around 1983, with a most probable origin in Portugal, but the tMRCAs of its subclusters are relatively recent, in 2008 and 2004, respectively. C_7 is related to a virus from Mozambique, a former Portuguese colony where subtype C is the predominant HIV-1 genetic form (Bellocchi et al., 2005).

It should be noted that 3 individuals in each of clusters F1_3 and C_7, composed entirely of men, declared being heterosexual, and that in clusters CRF02_1 (excluding the heterosexual-associated CRF02_1_3 subcluster) and A1_1, among sexually infected individuals the number of self-declared heterosexual men exceeds the number of women (8 vs. 3 in CRF02_1 and 6 vs. 3 in A1_1). This suggests that at least some of the self-declared heterosexual men within the clusters could in fact be MSM, who, due to social stigma and discrimination, do not declare their real sexual behaviors, as suggested by other authors who found similar discrepancies between self-reported heterosexual behavior and phylogenetic clustering with sequences from MSM (Hué et al., 2014; Hoenigl et al., 2016; Ragonnet-Cronin et al., 2018).

The descriptions of the four clusters here analyzed is in line with those of other non-subtype B clusters reported to have expanded among MSM in Western Europe. However there are some salient features of the clusters here described that should be highlighted: first, the relatively large size of CRF02_1_1 and A1_1, greater than most non-subtype B clusters reported among MSM in Western Europe; second, their wide geographic distribution among different countries, which contrasts with the predominantly within-country clustering found by other authors in Europe (Frentz et al., 2013; Paraskevis et al., 2019); and, third, the rapid expansion of F1_3, with 36 diagnoses in only 4 years.

The expansion of large clusters among MSM in recent years, as here reported, reflects the existence of high risk sexual behaviors, which should prompt public health authorities to implement public health measures aimed at preventing HIV-1 transmission in this population, including behavioral interventions to reduce risky practices, preexposure prophylaxis (Grant et al., 2010; Volk et al., 2015; McCormack et al., 2016),

and early diagnosis and treatment of HIV-1 infections (European Centre for Disease Prevention and Control, 2015; United Nations Population Fund, the Global Forum on MSM and HIV, United Nations Development Programme, World Health Organization, United States Agency for International Development, and World Bank, 2015). Prevention of HIV-1 transmission among MSM could also result in a reduction of heterosexually-transmitted infections, which can have their source in MSM networks, as seen in subcluster CRF02_1_3, associated with heterosexual transmission, and as reported by other authors (Oster et al., 2015; Esbjörnsson et al., 2016).

Continued HIV-1 molecular surveillance will be necessary to gain insight in real time on the dynamics of expansion of transmission networks, which will allow to focus prophylactic efforts in populations with the highest risk of HIV-1 acquisition and ongoing transmission (Little et al., 2014; Wang et al., 2015; Poon et al., 2016; Ratmann et al., 2016; Brenner et al., 2017; Chaillon et al., 2017; German et al., 2017; Oster et al., 2018; Wertheim et al., 2018) and to monitor the efficacy of public health interventions aimed at controlling the epidemic (Wertheim et al., 2011; Magiorkinis et al., 2018). HIV-1 molecular surveillance can also provide important information for the design of vaccine immunogens adapted to the major HIV-1 variants actively propagating in different areas, considering the correlation of HIV-1 clades to susceptibility to protective immune responses (Cao et al., 2000; Thomson et al., 2002; Binley et al., 2004; Geldmacher et al., 2007; Seaman et al., 2010; Hraber et al., 2014), and with potential to induce broadly neutralizing antibody responses (Kouyos et al., 2018), and will allow to obtain reagents derived from these variants for use in vaccine-related research (Cuevas et al., 2010; Revilla et al., 2011; Hora et al., 2016). These reagents will also be useful for studies on the biological basis of increased pathogenicity (Baeten et al., 2007; Kiwanuka et al., 2010; Li et al., 2014; Pérez-Álvarez et al., 2014; Kouri et al., 2015; Venner et al., 2016) and transmissibility (Kiwanuka et al., 2009) and diminished response to ARV drugs (Pernas et al., 2014; Cid-Silva et al., 2018) exhibited by some HIV-1 variants.

MEMBERS OF THE SPANISH GROUP FOR THE STUDY OF NEW HIV DIAGNOSES

Basque Country: Hospital Universitario de Basurto, Bilbao: Josefa Muñoz, María Carmen Nieto, María Zuriñe Zubero, Silvia Hernáez-Crespo; Hospital Universitario de Cruces, Bilbao: Luis Elorduy Otazua, Leyre López Soria, Koldo Agirrebengoa; Hospital de Galdakao: María José López de Goicoechea, José Mayo; Hospital Universitario Donostia, San Sebastián: Gustavo Cilla, Julio Arrizabalaga, José Antonio Iribarren, María Jesús Bustinduy, María Julia Echevarría. Hospital Universitario de Álava, Vitoria: María Jesús Lezaun, José Joaquín Portu, Carmen Gómez-González. **Galicia**: Complejo Hospitalario universitario de Ferrol, Ferrol, A Coruña: Ana Mariño, Patricia Ordóñez, Hortensia Álvarez, Nieves Valcarce; Complejo Hospitalario Universitario de A Coruña: Ángeles Cañizares, María Ángeles Castro. Hospital Universitario Lucus Augusti, Lugo: María Amparo Coira, María José López-Álvarez, Ramón Rabunal; Complejo Hospitalario Universitario de Ourense: Juan García-Costa, Ricardo Fernández-Rodríguez, Raúl Rodríguez-Pérez, Jorge Guitián; Complejo Hospitalario Universitario de Vigo, Vigo, Pontevedra: Antonio Ocampo, Celia Miralles, Sonia Pérez-Castro; Complejo Hospitalario de Pontevedra: Matilde Trigo, Julio Diz-Arén, María Ángeles Pallarés. **Navarra**: Complejo Hospitalario de Navarra, Pamplona: Carmen Ezpeleta Baquedano, Aitziber Aguinaga, María Gracia Ruiz de Alda. **Madrid**: Centro Sanitario Sandoval, Madrid: Jorge del Romero, Carmen Rodríguez, Mar Vera, Óscar Ayerdi; Hospital de Fuenlabrada: María Isabel García-Arata, Santiago Prieto-Menchero; Hospital Clínico Universitario San Carlos, Madrid: Esther Culebras, Iciar Rodríguez-Avial; Fundación Jiménez Díaz, Madrid: Raquel Téllez, Miguel Górgolas, Manuel Fernández-Guerrero, Olalla Calabia, Rosa García-Delgado; Hospital Severo Ochoa, Leganés: Sara María Quevedo, Lucía Puente, Manuel Álamo; Hospital Universitario Puerta de Hierro, Majadahonda: Alfonso Alfange, Sara de la Fuente. **Castilla y León**: Hospital Clínico Universitario de Valladolid: Carmen Hinojosa, Carlos Dueñas, Begoña Monteagudo, Edita Sánchez; Hospital Río Hortega, Valladolid: Carmen Ramos Sánchez, Pablo Bachiller, Helmuth Guillén; Hospital Virgen de la Concha, Zamora: Teresa Martínez-Domínguez, Rosa Martínez-González. **La Rioja**: Hospital San Pedro: José Ramón Blanco, Miriam Blasco. **Aragón**: Hospital Universitario Miguel Servet, Zaragoza: Ana María Martínez-Sapiña, Diego Ortega Larrea. **Castilla-La Mancha**: Hospital Virgen de la Salud, Toledo: César Gómez-Hernando, José Largo-Pau; **Comunitat Valenciana**: Hospital Universitari Sant Joan d'Alacant: Fernando Buñuel, Ana Infante.

DATA AVAILABILITY

The datasets generated for this study can be found in GenBank, MK177651–MK177824, MK177825–MK177829, KT276258, KY496622, and KY989952.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee' with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Bioethics and Animal Well-being Committee of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain.

AUTHOR CONTRIBUTIONS

MT, ED, and LP-Á conceived the study and supervised the experimental work. ED, MT, MC, AF-G, FD-F, JC, JM-L, and MS processed sequences and performed phylogenetic analyses.

MT and FD-F performed phylodynamic analyses. HG performed data curation and phylogenetic analyses. SB, VM, AF-G, MS-M, EG-B, and CC performed experimental work. The members of the Spanish Group for the Study of New HIV Diagnoses recruited patients and obtained epidemiological data. MT wrote the manuscript with contributions from the other authors.

FUNDING

This work was funded through Acción Estratégica en Salud Intramural (AESI), Instituto de Salud Carlos III, project "Estudios sobre vigilancia epidemiológica molecular del VIH-1 en España," PI16CIII/00033; Red de Investigación en SIDA (RIS), Instituto de Salud Carlos III, Subdirección General de Evaluación y Fondo Europeo de Desarrollo Regional (FEDER), Plan Nacional I+D+I, project RD16ISCIII/0002/0004; scientific agreements with Consellería de Sanidade, Government of Galicia (MVI 1004/16) and Osakidetza-Servicio Vasco de Salud, Government of Basque Country (MVI 1001/16); European Research Infrastructures for Poverty Related Diseases (EURIPRED). Seventh Framework Program: FP7-Capacities-infrastructures-2012-1, grant agreement 312661; and Dirección General de Farmacia, Ministerio de Sanidad, Servicios Sociales e Igualdad, Government of Spain (grant EC11-272).

ACKNOWLEDGMENTS

We would like to thank José Antonio Taboada, from Consellería de Sanidade, Xunta de Galicia, and Daniel Zulaika, from Osakidetza-Servicio Vasco de Salud, for their support of this study, and the personnel at the Genomic Unit, Instituto de Salud Carlos III, for technical assistance in sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00655/full#supplementary-material

REFERENCES

Abecasis, A. B., Martins, A., Costa, I., Carvalho, A. P., Diogo, I., Gomes, P., et al. (2011). Molecular epidemiological analysis of paired pol/env sequences from Portuguese HIV type 1 patients. *AIDS Res. Hum. Retroviruses* 27, 803–805. doi: 10.1089/AID.2010.0312

Abecasis, A. B., Wensing, A. M., Paraskevis, D., Vercauteren, J., Theys, K., van de Vijver, D. A., et al. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 10:7. doi: 10.1186/1742-4690-10-17

Ambrosioni, J., Junier, T., Delhumeau, C., Calmy, A., Hirschel, B., Zdobnov, E., et al. (2012). Impact of highly active antiretroviral therapy on the molecular epidemiology of newly diagnosed HIV infections. *AIDS* 26, 2079–2086. doi: 10.1097/QAD.0b013e32835805b6

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Apetrei, C., Loussert-Ajaka, I., Collin, G., Letourneur, F., Duca, M., Saragosti, S., et al. (1997). HIV type 1 subtype F sequences in Romanian children and adults. *AIDS Res. Hum. Retroviruses* 13, 363–365. doi: 10.1089/aid.1997.13.363

Audelin, A. M., Cowan, S. A., Obel, N., Nielsen, C., Jorgensen, L. B., and Gerstoft, J. (2013). Phylogenetics of the Danish HIV epidemic: the role of very late presenters in sustaining the epidemic. *J. Acquir. Immune Defic. Syndr.* 62, 102–108. doi: 10.1097/QAI.0b013e318276becc

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., et al. (2007). HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.* 195, 1177–1180. doi: 10.1086/512682

Bandea, C. I., Ramos, A., Pieniazek, D., Pascu, R., Tanuri, A., Schochetman, G., et al. (1995). Epidemiologic and evolutionary relationships between Romanian and Brazilian HIV-subtype F strains. *Emerg. Infect. Dis.* 1, 91–93. doi: 10.3201/eid0103.950305

Bartolo, I., Abecasis, A. B., Borrego, P., Barroso, H., McCutchan, F., Gomes, P., et al. (2011). Origin and epidemiological history of HIV-1 CRF14_BG. *PLoS One* 6:e24130. doi: 10.1371/journal.pone.0024130

Baryshev, P. B., Bogachev, V. V., and Gashnikova, N. M. (2014). HIV-1 genetic diversity in Russia: CRF63_02A1, a new HIV type 1 genetic variant spreading in Siberia. *AIDS Res. Hum. Retroviruses* 30, 592–597. doi: 10.1089/AID.2013.0196

Bello, G., Afonso, J. M., and Morgado, M. G. (2012). Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect. Genet. Evol.* 12, 1079–1086. doi: 10.1016/j.meegid.2012.03.014

Bellocchi, M. C., Forbici, F., Palombi, L., Gori, C., Coelho, E., Svicher, V., et al. (2005). Subtype analysis and mutations to antiviral drugs in HIV-1-infected patients from Mozambique before initiation of antiretroviral therapy: results from the DREAM programme. *J. Med. Virol.* 76, 452–458. doi: 10.1002/jmv.20382

Beloukas, A., Psarris, A., Giannelou, P., Kostaki, E., Hatzakis, A., and Paraskevis, D. (2016). Molecular epidemiology of HIV-1 infection in Europe: an overview. *Infect. Genet. Evol.* 46, 180–189. doi: 10.1016/j.meegid.2016.06.033

Beyrer, C., Baral, S. D., van Griensven, F., Goodreau, S. M., Chariyalertsak, S., Wirtz, A. L., et al. (2012). Global epidemiology of HIV infection in men who have sex with men. *Lancet* 380, 367–377. doi: 10.1016/S0140-6736(12)60821-6

Bezemer, D., de Wolf, F., Boerlijst, M. C., van Sighem, A., Hollingsworth, T. D., Prins, M., et al. (2008). A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. *AIDS* 22, 1071–1077. doi: 10.1097/QAD.0b013e3282fd167c

Bezemer, D., van Sighem, A., Lukashov, V. V., van der Hoek, L., Back, N., Schuurman, R., et al. (2010). Transmission networks of HIV-1 among men having sex with men in the Netherlands. *AIDS* 24, 271–282. doi: 10.1097/QAD.0b013e328333ddee

Binley, J. M., Wrin, T., Korber, B., Zwick, M. B., Wang, M., Chappey, C., et al. (2004). Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J. Virol.* 78, 13232–13252. doi: 10.1128/JVI.78.23.13232-13252.2004

Bobkova, M. (2013). Current status of HIV-1 diversity and drug resistance monitoring in the former USSR. *AIDS Rev.* 15, 204–212.

Bracho, M. A., Sentandreu, V., Alastrué, I., Belda, J., Juan, A., Fernández-García, E., et al. (2014). Emerging trends in CRF02_AG variants transmission among men who have sex with men in Spain. *J. Acquir. Immune Defic. Syndr.* 65, e130–e133. doi: 10.1097/01.qai.0000435602.73469.56

Brand, D., Moreau, A., Cazein, F., Lot, F., Pillonel, J., Brunet, S., et al. (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, 4010–4016. doi: 10.1128/JCM.01141-14

Brenner, B. G., Ibanescu, R. I., Hardy, I., and Roger, M. (2017). Genotypic and phylogenetic insights on prevention of the spread of HIV-1 and drug resistance in "real-world" Settings. *Viruses* 10:E10. doi: 10.3390/v10010010

Cao, H., Mani, I., Vincent, R., Mugerwa, R., Mugyenyi, P., Kanki, P., et al. (2000). Cellular immunity to human immunodeficiency virus type 1 (HIV-1) clades:

relevance to HIV-1 vaccine trials in Uganda. *J. Infect. Dis.* 182, 1350–1356. doi: 10.1086/315868

Carr, J. K., Nadai, Y., Eyzaguirre, L., Saad, M. D., Khakimov, M. M., Yakubov, S. K., et al. (2005). Outbreak of a West African recombinant of HIV-1 in Tashkent, Uzbekistan. *J. Acquir. Immune Defic. Syndr.* 39, 570–575.

Carvalho, A., Costa, P., Triunfante, V., Branca, F., Rodrigues, F., Santos, C. L., et al. (2015). Analysis of a local HIV-1 epidemic in Portugal highlights established transmission of non-B and non-G subtypes. *J. Clin. Microbiol.* 53, 1506–1514. doi: 10.1128/JCM.03611-14

Casado, C., Urtasun, I., Saragosti, S., Chaix, M. L., de Rossi, A., Cattelan, A. M., et al. (2000). Different distribution of HIV type 1 genetic variants in European patients with distinct risk practices. *AIDS Res. Hum. Retroviruses* 16, 299–304. doi: 10.1089/088922200309403

Castro, E., Khonkarly, M., Ciuffreda, D., Burgisser, P., Cavassini, M., Yerly, S., et al. (2010). HIV-1 drug resistance transmission networks in southwest Switzerland. *AIDS Res. Hum. Retroviruses* 26, 1233–1238. doi: 10.1089/aid.2010.0083

Chaillon, A., Essat, A., Frange, P., Smith, D. M., Delaugerre, C., Barin, F., et al. (2017). Spatiotemporal dynamics of HIV-1 transmission in France (1999-2014) and impact of targeted prevention strategies. *Retrovirology* 14:15. doi: 10.1186/s12977-017-0339-4

Chalmet, K., Staelens, D., Blot, S., Dinakis, S., Pelgrom, J., Plum, J., et al. (2010). Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infect. Dis.* 10:262. doi: 10.1186/1471-2334-10-262

Ciccozzi, M., Gori, C., Boros, S., Ruiz-Alvarez, M. J., Harxhi, A., Dervishi, M., et al. (2005). Molecular diversity of HIV in Albania. *J. Infect. Dis.* 192, 475–479. doi: 10.1086/431599

Cid-Silva, P., Margusino-Framinan, L., Balboa-Barreiro, V., Martín-Herranz, I., Castro-Iglesias, A., Pernas-Souto, B., et al. (2018). Initial treatment response among HIV subtype F infected patients who started antiretroviral therapy based on integrase inhibitors. *AIDS* 32, 121–125. doi: 10.1097/QAD.0000000000001679

Cuevas, M. T., Fernández-García, A., Pinilla, M., García-Álvarez, V., Thomson, M., Delgado, E., et al. (2010). Short communication: biological and genetic characterization of HIV type 1 subtype B and nonsubtype B transmitted viruses: usefulness for vaccine candidate assessment. *AIDS Res. Hum. Retroviruses* 26, 1019–1025. doi: 10.1089/aid.2010.0018

Cuevas, M. T., Muñoz-Nieto, M., Thomson, M. M., Delgado, E., Iribarren, J. A., Cilla, G., et al. (2009). HIV-1 transmission cluster with T215D revertant mutation among newly diagnosed patients from the Basque Country, Spain. *J. Acquir. Immune Defic. Syndr.* 51, 99–103. doi: 10.1097/QAI.0b013e318199063e

Dauwe, K., Mortier, V., Schauvliege, M., Van Den Heuvel, A., Fransen, K., Servais, J. Y., et al. (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.* 15:524. doi: 10.1186/s12879-015-1217-0

de Oliveira, T., Pillay, D., and Gifford, R. J. (2010). The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. *PLoS One* 5:e9311. doi: 10.1371/journal.pone.0009311

Delatorre, E. O., Bello, G., Eyer-Silva, W. A., Chequer-Fernandez, S. L., Morgado, M. G., and 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, 1369–1372. doi: 10.1089/aid.2011.0381

Delgado, E., Cuevas, M. T., Domínguez, F., Vega, Y., Cabello, M., Fernández-García, A., et al. (2015). Phylogeny and phylogeography of a recent HIV-1 subtype F outbreak among men who have sex with men in Spain deriving from a cluster with a wide geographic circulation in Western Europe. *PLoS One* 10:e0143325. doi: 10.1371/journal.pone.0143325

Delgado, E., Fernández-García, A., Vega, Y., Cuevas, T., Pinilla, M., García, V., et al. (2012). Evaluation of genotypic tropism prediction tests compared with *in vitro* co-receptor usage in HIV-1 primary isolates of diverse subtypes. *J. Antimicrob. Chemother.* 67, 25–31. doi: 10.1093/jac/dkr438

Delgado, E., Thomson, M. M., Villahermosa, M. L., Sierra, M., Ocampo, A., Miralles, C., et al. (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* 29, 536–543. doi: 10.1097/00126334-200204150-00016

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Dumitrescu, O., Kalish, M. L., Kliks, S. C., Bandea, C. I., and Levy, J. A. (1994). Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J. Infect. Dis.* 169, 281–288. doi: 10.1093/infdis/169.2.281

El Moussi, A., Thomson, M. M., Delgado, E., Cuevas, M. T., Nasr, M., Abid, S., et al. (2017). Genetic diversity of HIV-1 in Tunisia. *AIDS Res. Hum. Retroviruses* 33, 77–81. doi: 10.1089/AID.2016.0164

Esbjörnsson, J., Mild, M., Audelin, A., Fonager, J., Skar, H., Bruun, J. L., et al. (2016). HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. *Virus Evol.* 2:vew010. doi: 10.1093/ve/vew010

Esteves, A., Parreira, R., Piedade, J., Venenno, T., Franco, M., Germano, et al. (2003). Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Res. Hum. Retroviruses* 19, 511–517. doi: 10.1089/088922203766774568

Esteves, A., Parreira, R., Venenno, T., Franco, M., Piedade, J., Germano, et al. (2002). Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Res. Hum. Retroviruses* 18, 313–325. doi: 10.1089/088922202753519089

European Centre for Disease Prevention and Control (2015). *ECDC Guidance: HIV and STI Prevention Among Men Who Have Sex with Men*. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/hiv-sti-prevention-among-men-who-have-sex-with-men-guidance.pdf (accessed November 18, 2018).

European Centre for Disease Prevention and Control (2017). *HIV/AIDS Surveillance in Europe*. Available at: https://ecdc.europa.eu/en/publications-data/hivaids-surveillance-europe-2017-2016-data (accessed November 18, 2018).

Eyzaguirre, L. M., Erasilova, I. B., Nadai, Y., Saad, M. D., Kovtunenko, N. G., Gomatos, P. J., et al. (2007). Genetic characterization of HIV-1 strains circulating in Kazakhstan. *J. Acquir. Immune Defic. Syndr.* 46, 19–23. doi: 10.1097/QAI.0b013e318073c620

Fabeni, L., Alteri, C., Orchi, N., Gori, C., Bertoli, A., Forbici, F., et al. (2015). Recent transmission clustering of HIV-1 C and CRF17_BF strains characterized by NNRTI-related mutations among newly diagnosed men in central Italy. *PLoS One* 10:e0135325. doi: 10.1371/journal.pone.0135325

Fisher, M., Pao, D., Brown, A. E., Sudarshi, D., Gill, O. N., Cane, P., et al. (2010). Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. *AIDS* 24, 1739–1747. doi: 10.1097/QAD.0b013e32833ac9e6

Foley, B. T., Leitner, T., Paraskevis, D., and Peeters, M. (2016). Primate immunodeficiency virus classification and nomenclature: review. *Infect. Genet. Evol.* 46, 150–158. doi: 10.1016/j.meegid.2016.10.018

Foster, G. M., Ambrose, J. C., Hué, S., Delpech, V. C., Fearnhill, E., Abecasis, A. B., et al. (2014). Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of Circulating Recombinant Form (CRF) 50_A1D. *PLoS One* 9:e83337. doi: 10.1371/journal.pone.0083337

Frange, P., Meyer, L., Deveau, C., Tran, L., Goujard, C., Ghosn, J., et al. (2012). Recent HIV-1 infection contributes to the viral diffusion over the French territory with a recent increasing frequency. *PLoS One* 7:e31695. doi: 10.1371/journal.pone.0031695

Fransen, K., Buve, A., Nkengasong, J. N., Laga, M., and van der Groen, G. (1996). Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. *Lancet* 347:1403. doi: 10.1016/S0140-6736(96)91042-9

Frentz, D., Wensing, A. M., Albert, J., Paraskevis, D., Abecasis, A. B., Hamouda, O., et al. (2013). Limited cross-border infections in patients newly diagnosed with HIV in Europe. *Retrovirology* 10:36. doi: 10.1186/1742-4690-10-36

Geldmacher, C., Currier, J. R., Gerhardt, M., Haule, A., Maboko, L., Birx, D., et al. (2007). In a mixed subtype epidemic, the HIV-1 Gag-specific T-cell response is biased towards the infecting subtype. *AIDS* 21, 135–143. doi: 10.1097/01.aids.0000247589.77061.f7

German, D., Grabowski, M. K., and Beyrer, C. (2017). Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. *Sex Health* 14, 89–96. doi: 10.1071/SH16056

Gifford, R. J., de Oliveira, T., Rambaut, A., Pybus, O. G., Dunn, D., Vandamme, A. M., et al. (2007). Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. *J. Virol.* 81, 13050–13056. doi: 10.1128/JVI.00889-07

Gifford, R. J., Liu, T. F., Rhee, S. Y., Kiuchi, M., Hué, S., Pillay, D., et al. (2009). The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* 25, 1197–1198. doi: 10.1093/bioinformatics/btp134

Giuliani, M., Santoro, M. M., Lo, P. A., Cella, E., Scognamiglio, P., Lai, A., et al. (2013). Circulation of HIV-1 CRF02_AG among MSM population in central Italy: a molecular epidemiology-based study. *Biomed. Res. Int.* 2013:810617. doi: 10.1155/2013/810617

González-Domenech, C. M., Viciana, I., Delaye, L., Mayorga, M. L., Palacios, R., de la Torre, J., et al. (2018). Emergence as an outbreak of the HIV-1 CRF19_cpx variant in treatment-naive patients in southern Spain. *PLoS One* 13:e0190544. doi: 10.1371/journal.pone.0190544

Grant, R. M., Lama, J. R., Anderson, P. L., McMahan, V., Liu, A. Y., Vargas, L., et al. (2010). Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N. Engl. J. Med.* 363, 2587–2599. doi: 10.1056/NEJMoa1011205

Guimarães, M. L., Vicente, A. C., Otsuki, K., da Silva, R. F., Francisco, M., da Silva, F. G., et al. (2009). Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology* 6:39. doi: 10.1186/1742-4690-6-39

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and 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

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–W559. doi: 10.1093/nar/gki352

Hemelaar, J., Elangovan, H., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2018). Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9

HIV Sequence Database (2019). *HIV Sequence Database*. Available at https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html (accessed February 15, 2019).

Hoenigl, M., Chaillon, A., Kessler, H. H., Haas, B., Stelzl, E., Weninger, K., et al. (2016). Characterization of HIV transmission in South-East Austria. *PLoS One* 11:e0151478. doi: 10.1371/journal.pone.0151478

Hora, B., Keating, S. M., Chen, Y., Sanchez, A. M., Sabino, E., Hunt, G., et al. (2016). Genetic characterization of a panel of diverse HIV-1 isolates at seven international sites. *PLoS One* 11:e0157340. doi: 10.1371/journal.pone.0157340

Hraber, P., Korber, B. T., Lapedes, A. S., Bailer, R. T., Seaman, M. S., Gao, H., et al. (2014). Impact of clade, geography, and age of the epidemic on HIV-1 neutralization by antibodies. *J. Virol.* 88, 12623–12643. doi: 10.1128/JVI.01705-14

Hué, S., Brown, A. E., Ragonnet-Cronin, M., Lycett, S. J., Dunn, D. T., Fearnhill, E., et al. (2014). Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* 28, 1967–1975. doi: 10.1097/QAD.0000000000000383

Hué, S., Pillay, D., Clewley, J. P., and Pybus, O. G. (2005). Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4425–4429. doi: 10.1073/pnas.0407534102

Hughes, G. J., Fearnhill, E., Dunn, D., Lycett, S. J., Rambaut, A., and Leigh Brown, A. J. (2009). Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog.* 5:e1000590. doi: 10.1371/journal.ppat.1000590

Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062

Instituto Nacional de Estadística (2018). *Estadística del Padrón Continuo. Datos Provisionales a 1 de Enero de*. Available at: http://www.ine.es/jaxi/Datos.htm?path=/t20/e245/p04/provi/l0/&file=00000010.px (accessed November 18, 2018).

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kiwanuka, N., Laeyendecker, O., Quinn, T. C., Wawer, M. J., Shepherd, J., Robb, M., et al. (2009). HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS* 23, 2479–2484. doi: 10.1097/QAD.0b013e328330cc08

Kiwanuka, N., Robb, M., Laeyendecker, O., Kigozi, G., Wabwire-Mangen, F., Makumbi, F. E., et al. (2010). HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J. Acquir. Immune Defic. Syndr.* 54, 180–184. doi: 10.1097/QAI.0b013e3181c98fc0

Kouri, V., Khouri, R., Alemán, Y., Abrahantes, Y., Vercauteren, J., Pineda-Peña, A. C., et al. (2015). CRF19_cpx is an evolutionary fit HIV-1 variant strongly associated with rapid progression to AIDS in Cuba. *EBioMedicine* 2, 244–254. doi: 10.1016/j.ebiom.2015.01.015

Kouyos, R. D., Rusert, P., Kadelka, C., Huber, M., Marzel, A., Ebner, H., et al. (2018). Tracing HIV-1 strains that imprint broadly neutralizing antibody responses. *Nature* 561, 406–410. doi: 10.1038/s41586-018-0517-0

Kuiken, C., Thakallapalli, R., Esklid, A., and de Ronde, A. (2000). Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am. J. Epidemiol.* 152, 814–822. doi: 10.1093/aje/152.9.814

Lai, A., Bozzi, G., Franzetti, M., Binda, F., Simonetti, F. R., de Luca, A., et al. (2016). 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. *PLoS One* 11:e0146097. doi: 10.1371/journal.pone.0146097

Lai, A., Bozzi, G., Franzetti, M., Binda, F., Simonetti, F. R., Micheli, V., 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. doi: 10.1002/jmv.23891

Lai, A., Simonetti, F. R., Zehender, G., de Luca, A., Micheli, V., Meraviglia, P., 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. doi: 10.1371/journal.pone.0042223

Lapovok, I., Kazennova, E., Laga, V., Vasilyev, A., Utegenova, A., Abishev, A., et al. (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

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Leoz, M., Feyertag, F., Charpentier, C., Delaugerre, C., Wirden, M., Lemee, V., et al. (2013). Characterization of CRF56_cpx, a new circulating B/CRF02/G recombinant form identified in MSM in France. *AIDS* 27, 2309–2312. doi: 10.1097/QAD.0b013e3283632e0c

Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Leigh Brown, A. J. (2008). Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 5:e50. doi: 10.1371/journal.pmed.0050050

Li, Y., Han, Y., Xie, J., Gu, L., Li, W., Wang, H., et al. (2014). CRF01_AE subtype is associated with X4 tropism and fast HIV progression in Chinese patients infected through sexual transmission. *AIDS* 28, 521–530. doi: 10.1371/journal.pmed.0050050

Liitsola, K., Ristola, M., Holmstrom, P., Salminen, M., Brummer-Korvenkontio, H., Simola, S., et al. (2000). An outbreak of the circulating recombinant form AECM240 HIV-1 in the Finnish injection drug user population. *AIDS* 14, 2613–2615. doi: 10.1097/00002030-200011100-00028

Little, S. J., Kosakovsky Pond, S. L., Anderson, C. M., Young, J. A., Wertheim, J. O., Mehta, S. R., et al. (2014). Using HIV networks to inform real time prevention interventions. *PLoS One* 9:e98443. doi: 10.1371/journal.pone.0098443

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Louwagie, J., Delwart, E. L., Mullins, J. I., McCutchan, F. E., Eddy, G., and Burke, D. S. (1994). Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res. Hum. Retroviruses* 10, 561–567. doi: 10.1089/aid.1994.10.561

Lukashov, V. V., Kuiken, C. L., Vlahov, D., Coutinho, R. A., and Goudsmit, J. (1996). Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. *AIDS Res. Hum. Retroviruses* 12, 1179–1183. doi: 10.1089/aid.1996.12.1179

Magiorkinis, G., Karamitros, T., Vasylyeva, T. I., Williams, L. D., Mbisa, J. L., Hatzakis, A., et al. (2018). An innovative study design to assess the community effect of interventions to mitigate HIV epidemics using transmission-chain phylodynamics. *Am. J. Epidemiol.* doi: 10.1093/aje/kwy160 [Epub ahead of print].

McCormack, S., Dunn, D. T., Desai, M., Dolling, D. I., Gafos, M., Gilson, R., et al. (2016). Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* 387, 53–60. doi: 10.1016/S0140-6736(15)00056-2

Meloni, S. T., Kim, B., Sankale, J. L., Hamel, D. J., Tovanabutra, S., Mboup, 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. doi: 10.1128/JVI.78.22.12438-12445.2004

Monno, L., Brindicci, G., Lai, A., Punzi, G., Altamura, M., Simonetti, F. R., et al. (2012). An outbreak of HIV-1 BC recombinants in Southern Italy. *J. Clin. Virol.* 55, 370–373. doi: 10.1016/j.jcv.2012.08.014

Montavon, C., Toure-Kane, C., Liegeois, F., Mpoudi, E., Bourgeois, A., Vergne, L., et al. (2000). Most env and gag subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG. *J. Acquir. Immune Defic. Syndr.* 23, 363–374. doi: 10.1097/00126334-200004150-00001

Moskaleychik, F. F., Laga, V. Y., Delgado, E., Vega, Y., Fernández-García, A., Pérez, A., et al. (2015). [Rapid spread of the HIV-1 circular recombinant CRF02-AG in Russia and neighboring countries]. *Vopr. Virusol.* 60, 14–19.

Niculescu, I., Paraschiv, S., Paraskevis, D., Abagiu, A., Batan, I., Banica, L., et al. (2015). Recent HIV-1 outbreak among intravenous drug users in Romania: evidence for cocirculation of CRF14_BG and subtype F1 Strains. *AIDS Res. Hum. Retroviruses* 31, 488–495. doi: 10.1089/aid.2014.0189

Núñez, O., Hernando, V., and Díaz, A. (2018). Estimating the number of people living with HIV and the undiagnosed fraction in Spain in 2013. *AIDS* 32, 2573–2581. doi: 10.1097/QAD.0000000000001989

Op de Coul, E., van den Burg, R., Asjö, B., Goudsmit, J., Cupsa, A., et al. (2000). Genetic evidence of multiple transmissions of HIV type 1 subtype F within Romania from adult blood donors to children. *AIDS Res. Hum. Retroviruses* 16, 327–336. doi: 10.1089/088922200309205

Op de Coul, E. L., Lukashov, V. V., van Doornum, G. J., Goudsmit, J., and Coutinho, R. A. (1998). Multiple HIV-1 subtypes present amongst heterosexuals in Amsterdam 1988-1996: no evidence for spread of non-B subtypes. *AIDS* 12, 1253–1255. doi: 10.1097/00002030-199810000-00024

Oster, A. M., France, A. M., Panneer, N., Bañez Ocfemia, M. C., Campbell, E., Dasgupta, S., et al. (2018). Identifying clusters of recent and rapid HIV transmission through analysis of molecular surveillance data. *J. Acquir. Immune Defic. Syndr.* 79, 543–550. doi: 10.1097/QAI.0000000000001856

Oster, A. M., Wertheim, J. O., Hernandez, A. L., Bañez Ocfemia, M. C., Saduvala, N., and Hall, H. I. (2015). Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. *J. Acquir. Immune Defic. Syndr.* 70, 444–451. doi: 10.1097/QAI.0000000000000809

Palma, A. C., Araújo, F., Duque, V., Borges, F., Paixão, M. T., and Camacho, R. (2007). Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.* 7, 391–398. doi: 10.1016/j.meegid.2007.01.009

Pandey, S. S., Cherian, S., Thakar, M., and Paranjape, R. S. (2016). Short communication: phylogenetic and molecular characterization of six full-length HIV-1 genomes from India reveals a monophyletic lineage of Indian sub-subtype A1. *AIDS Res. Hum. Retroviruses* 32, 489–502. doi: 10.1089/AID.2015.0207

Paraskevis, D., Beloukas, A., Stasinos, K., Pantazis, N., de Mendoza, C., and Bannert, N. (2019). HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. *BMC Med.* 17:4. doi: 10.1186/s12916-018-1241-1

Paraskevis, D., Magiorkinis, E., Magiorkinis, G., Sypsa, V., Paparizos, V., Lazanas, M., et al. (2007). Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. *J. Infect. Dis.* 196, 1167–1176. doi: 10.1086/521677

Parczewski, M., Leszczyszyn-Pynka, M., Witak-Jedra, M., Szetela, B., Gasiorowski, J., Knysz, B., et al. (2017). Expanding HIV-1 subtype B transmission networks among men who have sex with men in Poland. *PLoS One* 12:e0172473. doi: 10.1371/journal.pone.0172473

Patiño-Galindo, J. A., Torres-Puente, M., Bracho, M. A., Alastrué, I., Juan, A., Navarro, D., et al. (2017). The molecular epidemiology of HIV-1 in the Comunidad Valenciana (Spain): analysis of transmission clusters. *Sci. Rep.* 7:11584. doi: 10.1038/s41598-017-10286-1

Patiño-Galindo, J. A., Torres-Puente, M., Gimeno, C., Ortega, E., Navarro, D., Galindo, M. J., et al. (2015). Expansion of the CRF19_cpx variant in Spain. *J. Clin. Virol.* 69, 146–149. doi: 10.1016/j.jcv.2015.06.094

Pérez-Álvarez, L., Delgado, E., Vega, Y., Montero, V., Cuevas, T., Fernández-García, A., et al. (2014). Predominance of CXCR4 tropism in HIV-1 CRF14_BG strains from newly diagnosed infections. *J. Antimicrob. Chemother.* 69, 246–253. doi: 10.1093/jac/dkt305

Pérez-Parra, S., Álvarez, M., Fernández-Caballero, J. A., Pérez, A. B., Santos, J., Bisbal, O., et al. (2018). Continued propagation of the CRF19_cpx variant among HIV-positive MSM patients in Spain. *J. Antimicrob. Chemother.* 73, 1031–1038. doi: 10.1093/jac/dkx474

Pernas, B., Grandal, M., Mena, A., Castro-Iglesias, A., Cañizares, A., Wyles, et al. (2014). High prevalence of subtype F in newly diagnosed HIV-1 persons in northwest Spain and evidence for impaired treatment response. *AIDS* 28, 1837–1840. doi: 10.1097/QAD.0000000000000326

Pineda-Peña, A. C., Theys, K., Stylianou, D. C., Demetriades, I., Abecasis, A. B., and Kostrikis, L. G. (2018). HIV-1 Infection in Cyprus, the Eastern Mediterranean European frontier: a densely sampled transmission dynamics analysis from 1986 to 2012. *Sci. Rep.* 8:1702. doi: 10.1038/s41598-017-19080-5

Poon, A. F., Gustafson, R., Daly, P., Zerr, L., Demlow, S. E., Wong, J., et al. (2016). Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV* 3, e231–e238. doi: 10.1016/S2352-3018(16)00046-1

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Ragonnet-Cronin, M., Hué, S., Hodcroft, E. B., Tostevin, A., Dunn, D., Fawcett, T., et al. (2018). Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV* 5, e309–e316. doi: 10.1016/S2352-3018(18)30062-6

Ragonnet-Cronin, M., Lycett, S. J., Hodcroft, E. B., Hué, S., Fearnhill, E., Brown, A. E., et al. (2016). Transmission of non-B HIV subtypes in the United Kingdom is increasingly driven by large non-heterosexual transmission clusters. *J. Infect. Dis.* 213, 1410–1418. doi: 10.1093/infdis/jiv758

Rambaut, A., Lam, T. T., Max, C. L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Ratmann, O., van Sighem, A., Bezemer, D., Gavryushkina, A., Jurriaans, S., Wensing, A., et al. (2016). Sources of HIV infection among men having sex with men and implications for prevention. *Sci. Transl. Med.* 8:320ra2. doi: 10.1126/scitranslmed.aad1863

Reis, M. N. G, Guimarñes, M. L., Bello, G., and Stefani, M. M. A. (2019). Identification of new HIV-1 circulating recombinant forms CRF81_cpx and CRF99_BF1 in central western brazil and of unique BF1 recombinant forms. *Front. Microbiol.* 10:97. doi: 10.3389/fmicb.2019.00097

Revilla, A., Delgado, E., Christian, E. C., Dalrymple, J., Vega, Y., Carrera, C., et al. (2011). Construction and phenotypic characterization of HIV type 1 functional envelope clones of subtypes G and F. *AIDS Res. Hum. Retroviruses* 27, 889–901. doi: 10.1089/AID.2010.0177

Seaman, M. S., Janes, H., Hawkins, N., Grandpre, L. E., Devoy, C., Giri, A., et al. (2010). Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* 84, 1439–1452. doi: 10.1128/JVI.02108-09

Shapiro, B., Rambaut, A., and Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* 23, 7–9. doi: 10.1093/molbev/msj021

Shcherbakova, N. S., Shalamova, L. A., Delgado, E., Fernández-García, A., Vega, Y., Karpenko, L. I., et al. (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, 912–919. doi: 10.1089/AID.2014.0075

Sierra, M., Thomson, M. M., Ríos, M., Casado, G., de Castro, R. O., Delgado, E., et al. (2005). The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. *Infect. Genet. Evol.* 5, 209–217. doi: 10.1016/j.meegid.2004.07.010

Simonetti, F. R., Lai, A., Monno, L., Binda, F., Brindicci, G., Punzi, G., et al. (2014). Identification of a new HIV-1 BC circulating recombinant form (CRF60_BC) in Italian young men having sex with men. *Infect. Genet. Evol.* 23, 176–181. doi: 10.1016/j.meegid.2014.02.007

Tamalet, C., Ravaux, I., Moreau, J., Bregigeon, S., Tourres, C., Richet, H., et al. (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.* 87, 1327–1333. doi: 10.1002/jmv.24184

Temereanca, A., Ene, L., Mehta, S., Manolescu, L., Duiculescu, D., and Ruta, S. (2013). Transmitted HIV drug resistance in treatment-naive Romanian patients. *J. Med. Virol.* 85, 1139–1147. doi: 10.1002/jmv.23572

Thomson, M. M., Delgado, E., Manjón, N., Ocampo, A., Villahermosa, M. L., Mariño, A., et al. (2001). HIV-1 genetic diversity in Galicia Spain: BG intersubtype recombinant viruses circulating among injecting drug users. *AIDS* 15, 509–516. doi: 10.1097/00002030-200103090-00010

Thomson, M. M., Fernández-García, A., Delgado, E., Vega, Y., Díez-Fuertes, F., Sánchez-Martínez, M., et al. (2012). Rapid expansion of a HIV-1 subtype F cluster of recent origin among men who have sex with men in Galicia, Spain. *J. Acquir. Immune Defic. Syndr.* 59, e49–e51. doi: 10.1097/QAI.0b013e3182400fc4

Thomson, M. M., and Nájera, R. (2001). Travel and the introduction of human immunodeficiency virus type 1 non-B subtype genetic forms into Western countries. *Clin. Infect. Dis.* 32, 1732–1737. doi: 10.1086/320764

Thomson, M. M., Pérez-Álvarez, L., and Nájera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* 2, 461–471. doi: 10.1016/S1473-3099(02)00343-2

United Nations Population Fund, the Global Forum on MSM and HIV, United Nations Development Programme, World Health Organization, United States Agency for International Development, and World Bank (2015). *Implementing Comprehensive HIV and STI Programmes with Men Who Have Sex with Men*. Available at: https://www.unfpa.org/es/node/13155

Vega, Y., Delgado, E., Fernández-García, A., Cuevas, M. T., Thomson, M. M., Montero, V., et al. (2015). Epidemiological surveillance of HIV-1 transmitted drug resistance in Spain in 2004-2012: relevance of transmission clusters in the propagation of resistance mutations. *PLoS One* 10:e0125699. doi: 10.1371/journal.pone.0125699

Venner, C. M., Nankya, I., Kyeyune, F., Demers, K., Kwok, C., Chen, P. L., et al. (2016). Infecting HIV-1 subtype predicts disease progression in women of sub-Saharan Africa. *EBioMedicine* 13, 305–314. doi: 10.1016/j.ebiom.2016.10.014

Verhofstede, C., Dauwe, K., Fransen, K., Van Laethem, K., Van den Wijngaert, S., Ruelle, J., et al. (2018). Phylogenetic analysis of the Belgian HIV-1 epidemic reveals that local transmission is almost exclusively driven by men having sex with men despite presence of large African migrant communities. *Infect. Genet. Evol.* 61, 36–44. doi: 10.1016/j.meegid.2018.03.002

Vinken, L., Fransen, K., Pineda-Peña, A. C., Alexiev, I., Balotta, C., Debaisieux, L., et al. (2017). HIV-1 sub-subtype F1 outbreak among MSM in Belgium. *Virus Evol.* 3(Suppl. 1):vew036.020. doi: 10.1093/ve/vew036.020

Volk, J. E., Marcus, J. L., Phengrasamy, T., Blechinger, D., Nguyen, D. P., Follansbee, S., et al. (2015). No new HIV infections with increasing use of HIV preexposure prophylaxis in a clinical practice setting. *Clin. Infect. Dis.* 61, 1601–1603. doi: 10.1093/cid/civ778

von Wyl, V., Kouyos, R. D., Yerly, S., Boni, J., Shah, C., Burgisser, P., et al. (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

Wang, X., Wu, Y., Mao, L., Xia, W., Zhang, W., Dai, L., et al. (2015). Targeting HIV prevention based on molecular epidemiology among deeply sampled subnetworks of men who have sex with men. *Clin. Infect. Dis.* 61, 1462–1468. doi: 10.1093/cid/civ526

Wertheim, J. O., Kosakovsky Pond, S. L., Little, S. J., and De Gruttola, V. (2011). Using HIV transmission networks to investigate community effects in HIV prevention trials. *PLoS One* 6:e27775. doi: 10.1371/journal.pone.0027775

Wertheim, J. O., Murrell, B., Mehta, S. R., Forgione, L. A., Kosakovsky Pond, S. L., Smith, D. M., et al. (2018). Growth of HIV-1 molecular transmission clusters in New York City. *J. Infect. Dis.* 218, 1943–1953. doi: 10.1093/infdis/jiy431

Yebra, G., de Mulder, M., Martín, L., Rodríguez, C., Labarga, P., Viciana, I., et al. (2012). Most HIV type 1 non-B infections in the Spanish cohort of antiretroviral treatment-naive HIV-infected patients (CoRIS) are due to recombinant viruses. *J. Clin. Microbiol.* 50, 407–413. doi: 10.1128/JCM.05798-11

Yebra, G., Frampton, D., Gallo Cassarino, T., Raffle, J., Hubb, J., Ferns, R. B., et al. (2018). A high HIV-1 strain variability in London, UK, revealed by full-genome analysis: results from the ICONIC project. *PLoS One* 13:e0192081. doi: 10.1371/journal.pone.0192081

Zehender, G., Ebranati, E., Lai, A., Santoro, M. M., Alteri, C., Giuliani, M., et al. (2010). Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J. Acquir. Immune Defic. Syndr.* 55, 156–160. doi: 10.1097/QAI.0b013e3181eb3002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 April 2019
doi: 10.3389/fmicb.2019.00749

Estimating HIV-1 Genetic Diversity in Brazil Through Next-Generation Sequencing

*Brunna M. Alves[1], Juliana D. Siqueira[1], Isabel M. Prellwitz[1], Ornella M. Botelho[1], Vanusa P. Da Hora[2], Sabri Sanabani[3], Patrícia Recordon-Pinson[4], Hervé Fleury[4], Esmeralda A. Soares[1] and Marcelo A. Soares[1,5]**

[1] Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, Brazil, [2] Laboratório de Biologia Molecular, Escola de Medicina, Universidade Federal do Rio Grande, Rio Grande do Sul, Brazil, [3] LIM-3, Hospital das Clinicas FMUSP, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, [4] CNRS MFP-UMR 5234, University Hospital of Bordeaux, University of Bordeaux, Bordeaux, France, [5] Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Approximately 36.7 million people were living with the human immunodeficiency virus (HIV) at the end of 2016 according to UNAIDS, representing a global prevalence rate of 0.8%. In Brazil, an HIV prevalence of 0.24% has been estimated, which represents approximately 830,000 individuals living with the virus. As a touristic and commercial hub in Latin America, Brazil harbors an elevated HIV genetic variability, further contributed by the selective pressure exerted by the host immune system and by antiretroviral treatment. Through the progress of the next-generation sequencing (NGS) techniques, it has been possible to expand the study of HIV genetic diversity, evolutionary, and epidemic processes, allowing the generation of HIV complete or near full-length genomes (NFLG) and improving the characterization of intra- and interhost diversity of viral populations. Greater sensitivity in the detection of viral recombinant forms represents one of the major improvements associated with this development. It is possible to identify unique or circulating recombinant forms using the near full-length viral genomes with increasing accuracy. It also permits the characterization of multiple viral infections within individual hosts. Previous Brazilian studies using NGS to analyze HIV diversity were able to identify several distinct unique and circulating recombinant forms and evidenced dual infections. These data unveiled unprecedented high rates of viral recombination and highlighted that novel recombinants are continually arising in the Brazilian epidemic. In the pooled analysis depicted in this report, HIV subtypes have been determined from HIV-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul). Combined data analysis showed a significant prevalence of recombinant forms (29%; 101/350), and a similar 26% when only NFLGs were considered. Moreover, the analysis was able to evidence the occurrence of multiple infections in some individuals. Our data highlight the great HIV genetic diversity found in Brazil and unveils a more accurate scenario of the HIV evolutionary dynamics in the region.

Keywords: HIV-1, genetic diversity, NGS, NFLG, subtype

OPEN ACCESS

Edited by:
Joris Hemelaar,
University of Oxford, United Kingdom

Reviewed by:
Frank van der Meer,
University of Calgary, Canada
Guido van Marle,
University of Calgary, Canada

***Correspondence:**
Marcelo A. Soares
masoares@inca.gov.br

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 October 2018*
Accepted: *25 March 2019*
Published: *09 April 2019*

Citation:
Alves BM, Siqueira JD, Prellwitz IM, Botelho OM, Da Hora VP, Sanabani S, Recordon-Pinson P, Fleury H, Soares EA and Soares MA (2019) Estimating HIV-1 Genetic Diversity in Brazil Through Next-Generation Sequencing. Front. Microbiol. 10:749. doi: 10.3389/fmicb.2019.00749

INTRODUCTION

The first registry of AIDS was reported in the beginning of the 1980s, and until now 77 million people have become infected with human immunodeficiency virus (HIV) and 35 million have died from AIDS-related causes. By 2017, it was estimated that 36.9 million people worldwide were living with HIV (UNAIDS, 2018). Harboring over one-third of the total population of Latin America, Brazil accounts for nearly half of the new HIV infections and of the estimated total of individuals living with HIV/AIDS (48 and 46%, respectively) in the region (UNAIDS, 2018). This scenario, along with the high error-prone rate of the viral reverse transcriptase (RT), high virus replication rates and recombination events, contributes to the remarkable accumulation of genetic diversity in its population during the course of infection, further influenced by selective pressure exerted by the host immune system and by antiretroviral treatment (Roberts et al., 1988; Overbaugh and Bangham, 2001; Zhuang et al., 2002; Santoro and Perno, 2013).

The surveillance of HIV diversity assists to monitor the emergence of new subtypes and the presence of novel strains in a given geographic location (Hemelaar, 2013). The great diversity of HIV-1 group M, which disseminated on a global scale and dominates the current AIDS pandemic, allowed the phylogenetic classification in nine pure subtypes (A–D, F–H, J, and K), sub-subtypes (A1–A5, F1–F2), circulating recombinant forms (CRFs) and unique recombinant forms (URFs) (Robertson et al., 2000). Currently, more than 90 CRFs have been reported in the HIV Sequence Database of the United States Los Alamos National Laboratory[1]. Recombinant viruses are the result of simultaneous infection by multiple viruses during a single transmission event (co-infection) or from sequential infection at multiple transmission events (superinfection) (Yerly et al., 2004). Molecular epidemiology studies show that the overall distribution of HIV-1 groups, subtypes and recombinant forms is highly heterogeneous, with significant differences in the size of the epidemic and the geographical distribution. Overall, subtype C is responsible for half of the current infections (48%), followed by subtypes A (12%), and B (11%). A high prevalence of recombinant forms, which account for at least 21% of HIV-1 infections worldwide, is also noteworthy (Hemelaar, 2012). As seen in Latin America and the Caribbean countries, subtype B prevails in most parts of Brazil, followed by subtypes F1, C, D, and diverse recombinant forms. Southern Brazil, however, presents a distinct epidemiological pattern, with a higher prevalence of subtypes C, B, and BC recombinants (Cardoso et al., 2009; Machado et al., 2009, 2017; de Medeiros et al., 2011; Almeida et al., 2012; Graf and Pinto, 2013; Velasco-de-Castro et al., 2014; Junqueira and Almeida, 2016; Pessoa et al., 2016; Delatorre et al., 2017; Filho and Brites, 2017; Lima et al., 2017).

Through the progress of next-generation sequencing (NGS) techniques it became possible to expand the study of HIV genetic diversity, evolutionary and epidemic processes, allowing the generation of HIV complete or near full-length genomes (NFLG) and improving the characterization of intra- and interhost diversity of viral populations. Greater sensitivity and accuracy in the detection of viral recombinant forms represents one of the major improvements associated with this development, since most of the previous studies were based on partial HIV genomic sequences, mainly within the *pol* gene region due to the interest in determining drug resistance mutational patterns, resulting in underestimation of the frequency of recombinant forms (Thomson and Najera, 2005; Hemelaar et al., 2011; Marques et al., 2018). It also permits the characterization of multiple viral infections within individual hosts. Previous Brazilian studies using NGS to analyze HIV diversity were able to identify several distinct unique and circulating recombinant forms and evidenced dual infections (Pessoa et al., 2014b, 2015, 2016; Alves et al., 2017; Marques et al., 2018). In the pooled analysis depicted in this report, we pooled publically available Brazilian sequences obtained by NGS and new genetic data from HIV-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul). Upon analyzing the HIV subtypes of this large cohort, our study was able to unveil, with unprecedented accuracy, high rates of viral recombination and highlighted that novel recombinants are continually arising in the Brazilian epidemic.

MATERIALS AND METHODS

Study Population and Sample Collection

A total of 84 convenience samples were used for generating the experimental data presented in this study. These were from HIV-1-seropositive patients recruited between February 2016 and December 2017 during the routine services conducted at Sexually Transmitted Diseases/HIV ambulatory at Hospital Federal de Ipanema (HFI) and at Hospital Universitário Clementino Fraga Filho (HU-UFRJ), both located in Rio de Janeiro, southeastern Brazil, and at Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU-FURG), located in Rio Grande, southern Brazil. Clinical and epidemiological data were obtained through a questionnaire and 10 ml of whole peripheral blood were collected. This research was approved by the Ethics Committees in Research of the Brazilian National Cancer Institute – INCA and of HFI (CAAE 52862016.9.0000.5274), HUCFF-UFRJ (CAAE 56604816.2.0000.5257), and HU-FURG (CAAE 52862016.9.3001.5324). The inclusion criteria were being 18 years or greater, being under first antiretroviral scheme and being upon virological success (undetectable HIV viral load) for the last 12 months. A fraction of this casuistic, 32 patients from HFI, has been previously described (Alves et al., 2017). We pooled these data with all Brazilian HIV-1 data comprising NFLG and partial sequences determined by NGS and publically available (Pessoa et al., 2015, 2016). All studies included in the present report (either experimentally determined herein or retrieved from the literature) used a similar methodology to amplify the HIV NFLG and sequence them in an Illumina MiSeq platform. Multiple infection analyses were also made by *de novo* assembly as described below.

[1] http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html

DNA Extraction and PCR of Proviral DNA

The patients' genomic DNA containing their HIV-1 proviruses was extracted from whole blood with the Genomic DNA Extraction Kit (Real Genomics, BioAmerica, Inc.) following manufacturer's instructions. Nested PCR performed using Platinum™ Taq DNA Polymerase High Fidelity (Life Technologies) was carried out in a Veriti® 96-Well Thermal Cycler (Life Technologies, Carlsbad, United States) for the amplification of HIV NFLGs. The strategies comprised the amplification of four to five fragments, from 2 to 3 kb each, spanning the whole HIV genome (Sanabani et al., 2006b; Ode et al., 2015). After visualization using GelRed (Biotium, Hayward, CA, United States) in 1% agarose gels, duplicated independent PCR-positive products were pooled directly to avoid representativeness of PCR-based errors and their interpretation as minority variants in the population. PCR products were purified with the GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, MA, United States) and their concentration was measured in a NanoDrop ND 1000 apparatus (Thermo Fisher Scientific, MA, United States). The purified products were diluted to 4 ng/μL and pooled per patient.

Library Construction and NGS

Libraries were prepared with the Nextera XT DNA Sample Preparation kit (Illumina Inc., San Diego, United States) according to the manufacturer's protocol, except that the starting material was diluted to 0.4 ng/μL. The library construction consists of a fragmentation step using transposon technology, followed by a PCR step where dual indexes were added to the fragments. After this process, libraries were quantified by qPCR with the KAPA library quantification kit (Kapa Biosystems, MA, United States) or by fluorometric quantitation with the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, EUA). Libraries were diluted, and pooled prior to sequencing in a MiSeq Illumina platform (2 × 301 cycles paired-end run) (Illumina) with 1% denatured PhiX DNA as a sequencing control.

Data Analysis

The analysis of the obtained files was performed in Geneious v.9.1.3 using the same alignment parameters for a reference-based genome assembly described by Dudley et al. (2014). Reads were assembled using an annotated HIV-1 HXB2 reference sequence and 10 iterations to obtain each viral genome sequence. The presence of multiple infections was investigated by *de novo* assembly performed with IVA (*Iterative Virus Assembler*) with default parameters (Hunt et al., 2015). This assembler is based on seed sequences that are iteratively and conservatively extended into a contig using reads that have a perfect match. After this process, the program uses those contig as reference in a reference genome aligner to extend the initial contigs obtained (Hunt et al., 2015). The contig sequences obtained were submitted to a BLAST nucleotide analysis to discard those of human origin and confirmed with the Los Alamos National Laboratories HIV BLAST Tool[2]. All samples suggestive of multiple infections (more than one contig representing the same genomic region) were re-aligned using their respective IVA-assembled contigs as references in the Geneious program. More restrictive assembly parameters were used, such as allowing mapping of paired reads only when both mapped to the same contig and discarding reads that mapped simultaneously to more than one contig. To investigate if these contigs were derived from the same virus, sequences were submitted to hypermutation analysis using the Hypermut2.0 tool available at Los Alamos HIV Database. The contigs were considered hypermutated if the *p*-value was ≤ 0.05 when comparing the number of APOBEC G-to-A signature mutations with the control context. The overlapped region between the contigs obtained per sample was compared to all sequences publically available at the BLASTn Database[3]. The top 10 hits of each contig were retrieved from this database to construct phylogenetic trees and check its clustering profile.

Phylogenetic Analysis

The consensus sequence for each sample was extracted from the reference-guided assembly described above using the 50% stringency setting and classified using maximum likelihood phylogenetic analysis performed with PhyML v.3.0 and the best model of nucleotide substitution defined with Model Generator (Keane et al., 2006; Guindon et al., 2010). To investigate HIV-1 recombination the *bootscaning* tool of Simplot v.3.5.1 was used with the following parameters: *window* = 400 pb; *steps* = 40 pb; *T/t* = 2.0; *gapstrip* = on; *replicas* = 100; *nucleotide substitution model* = F84; *method* = Maximum Likelihood (Lole et al., 1999). Phylogenetic analyses were repeated for recombinant sequences considering the *bootscanning* breakpoint analysis (data not shown). The sequences obtained in this study were submitted to the GenBank under the accession numbers MK041550-MK041589. The raw sequencing reads were deposited to the Sequence Read Archive (SRA) under the numbers SRR7993842-SRR7993872.

RESULTS

In this study, we included data previously published by our group from patients followed-up at HFI and all HIV-1 sequences obtained by NGS publicly available. A detailed description of the studied populations can be found in their respective articles (Pessoa et al., 2015, 2016; Alves et al., 2017). We focused on the HIV-1 subtype classification, identification of HIV recombinants and multiple infection investigation for our novel patients. As previous published, the patients from HFI were mostly males (75%) with a median age of 38 years at the time of sample collection. Regarding our new cohort, we also found a prevalence of males (67%) among the patients from HU-UFRJ with a median age of 43.5 years. Unlike the other centers, a greater number of female patients (62%) were observed among patients from HU-FURG, with a median age of 43 years. Clinical and epidemiological characteristics of the three cohorts are compiled in **Table 1**. Regarding antiretroviral treatment, 19 patients from HI (19/32, 59%),

[2]https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html

[3]https://blast.ncbi.nlm.nih.gov/

TABLE 1 | Demographic and clinical characteristics of HIV-positive patients.

Characteristic	HI (*n* = 32)	HU- UFRJ (*n* = 12)	HU-FURG (*n* = 40)
Males (%)	24 (75%)	8 (67%)	15 (38%)
Age (years)	38	43.5	43
(median; IQR_{50})	(31.5–45.25)	(41–52)	(35–49.5)
Median baseline CD4+ T-cell counts	712.5	1117[1]	780.5
(cells/mm^3; IQR_{50})	(606.5–856)		(680.75–938.5)
Median baseline CD8+ T-cell counts	657.5	NA	848
(cells/mm^3; IQR_{50})	(529–1047.25)		(750.5–1053)
Median time since HIV diagnosis	4.7	13.3	4.83
(years; IQR_{50})	(3.9–6.5)	(7.7–14.8)	(2.42–10.9)
Median time from HIV diagnosis to start of ARV treatment	1.2	1.7	0.88
(years; IQR_{50})	(0.6–2.8)[2]	(0.1–5.3)	(0.25–3.2)
Median treatment time	3.1	9.34	3.83
(years; IQR_{50})	(2.4–3.9)	(4.6–11.4)	(2.2–4.77)

[1] *Information available for only two patients (1201 and 1033 cells/mm^3).* [2] *Data based on 31 patients. NA, information not available. IQR_{50}, interquartile range.*

four patients from HU-UFRJ (4/12, 33%), and 23 patients from HU-FURG (23/40, 58%), were under the HAART scheme preconized by Brazilian Ministry of Healthy at the time of sample collection, composed of tenofovir (TDF), lamivudine (3TC), and efavirenz (EFV). All schemes used by the patients are described in **Table 2**.

Overall, we were successful in sequencing by NGS at least 2 of the 4/5 overlapping PCR fragments of 48 samples from our cohort included in this study (57%, 48/84). Of those, 28 samples (58%; 28/48) had the NFLG obtained. The remaining partial genome sequences had complete Gag CDS (coding sequence) for nine samples (45%, 9/20), Pol CDS for 4 (20%, 4/20), and

TABLE 2 | Distribution of subtypes and HAART regimen exposure across the 48 HIV-1 genome sequences analyzed.

Patient	Subtype/ URF	HAART Regimen	Patient	Subtype/ URF	HAART Regimen
1-HI	B	AZT+3TC+NVP	**31-HI**	**B**	TDF+3TC+EFV
2-HI[#]	**B**	AZT+3TC+EFV	**32-HI**	**BF**	TDF+3TC+EFV
3-HI	**B**	TDF+3TC+EFV	**2-HU-UFRJ**	**B**	TDF+3TC+EFV
4-HI	B	TDF+3TC+EFV	3-HU-UFRJ	B	AZT+3TC+EFV
5-HI	**B**	AZT+3TC+ATV	**4-HU-UFRJ**	**B**	AZT+3TC+EFV
6-HI	F1	AZT+3TC+ATV/r	**6-HU-UFRJ**	**B**	AZT+3TC+EFV
8-HI	BC	TDF+3TC+EFV	**7-HU-UFRJ**	**B**	AZT+3TC+EFV
11-HI	**BF**	TDF+3TC+EFV	**9-HU-UFRJ**	**F1**	AZT+3TC+AZT/r
12-HI	**B**	TDF+3TC+EFV	**10-HU-UFRJ**	**B**	TDF+3TC+EFV
13-HI	**B**	TDF+3TC+EFV	**12-HU-UFRJ**	**B**	TDF+3TC+EFV
14-HI	**B**	AZT+3TC+LPV/r	**1-HU-FURG**	**C**	TDF+3TC+EFV
15-HI	**B**	TDF+3TC+EFV	2-HU-FURG	C	TDF+3TC+EFV
16-HI	**B**	AZT+3TC+FPV/r	3-HU-FURG	BC	AZT+3TC+EFV
17-HI	**BF1**	AZT+3TC+ATV/r	**6-HU-FURG**	**C**	TDF+3TC+AZT/r
18-HI	**B**	TDF+3TC+EFV	**7-HU-FURG**	**C**	TDF+3TC+DRV/r
19-HI	**B**	AZT+3TC+EFV	**8-HU-FURG**	**C**	TDF+3TC+ATV
20-HI	B	TDF+3TC+EFV	10-HU-FURG	C	TDF+3TC+EFV
21-HI	**B**	TDF+EFV+FTC	12-HU-FURG	C	TDF+3TC+EFV
22-HI	**B**	TDF+3TC+EFV	13-HU-FURG	C	TDF+3TC+EFV
23-HI	**B**	TDF+3TC+EFV	14-HU-FURG	BF1	AZT+3TC+AZT/r
26-HI	B	TDF+3TC+EFV	16-HU-FURG	C	TDF+3TC+EFV
27-HI	B	TDF+3TC+EFV	17-HU-FURG	C	AZT+3TC+LPV/r
28-HI	BF	TDF+3TC+EFV	18-HU-FURG	BF1	TDF+3TC+EFV
29-HI	B	TDF+3TC+EFV	20-HU-FURG	BC	TDF+3TC+EFV

[#] *The near full-length genomes are represented in bold.*

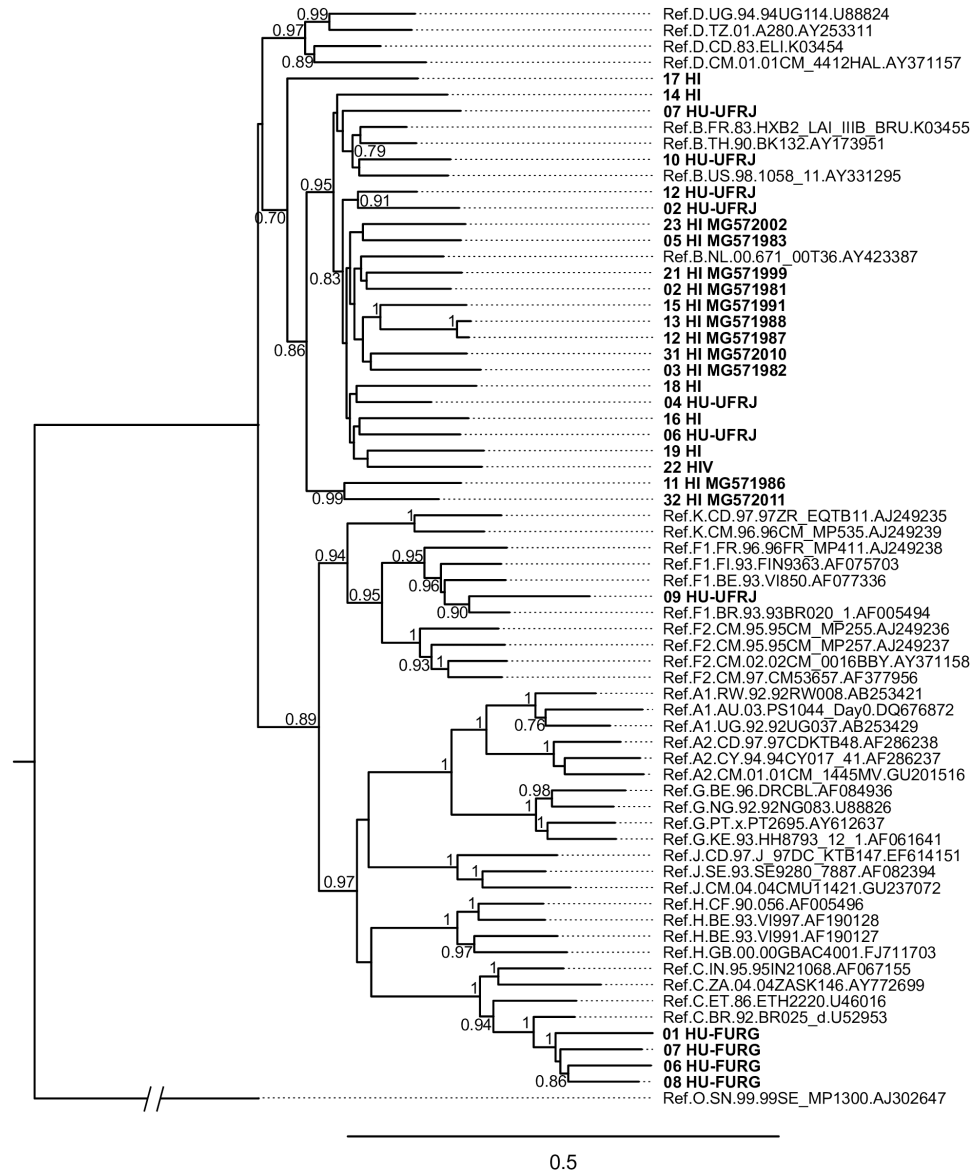


FIGURE 1 | Phylogenetic analysis of HIV near full-length genomes (NFLG) from this study. The maximum likelihood analysis was performed with 1,000 bootstrap iterations. The tree contains 28 HIV-1 proviral sequences obtained from Hospital Federal de Ipanema (HFI), Hospital Universitário Clementino Fraga Filho-UFRJ (HU-UFRJ), Rio de Janeiro, and Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU-FURG), Rio Grande do Sul (represented in bold) and reference sequences of HIV-1 subtypes (named by subtype, country, year, and GenBank accession number). HI sequences determined in a previous study (Alves et al., 2017) are named with their respective GenBank accession numbers. Only bootstrap values greater than 0.7 are shown.

Env CDS for 12 (60%, 12/20). Six samples from HI, 4 from HU-UFRJ, and 26 from HU-FURG (43%) failed to have more than one viral DNA fragment PCR-amplified and were excluded from further analyses. Of the 302 NGS Brazilian sequences previously available in the literature, which information was also included in this study, 247 were NFLG and 55 were partial sequences (Pessoa et al., 2015, 2016).

HIV-1 consensus sequences were subjected to phylogenetic analysis to determine their subtype/CRF classification. The NFLGs obtained by our group were mostly classified as HIV-1 subtype B (71%; 20/28), followed by subtype C (14%; 4/28), recombinant forms (11%; 3/28), and subtype F1 (4%; 1/28) (**Figure 1**). **Table 2** describes the subtype classification of each sample included. The recombinants were classified as distinct URFs involving subtypes B and F1 based on Simplot analysis (**Figure 2**). Two of them, HI-11 and HI-32, were already described in our former study (Alves et al., 2017). The sequences of the 20 HIV-1 partial genomes were predominantly subtype B (7/20), C (6/20), and recombinants forms (6/20) (**Table 2**). A single subtype F1 sample was identified. With respect to the recombinants found in our study and submitted to Simplot analysis, two were identical URF_BF1 and a third one was a

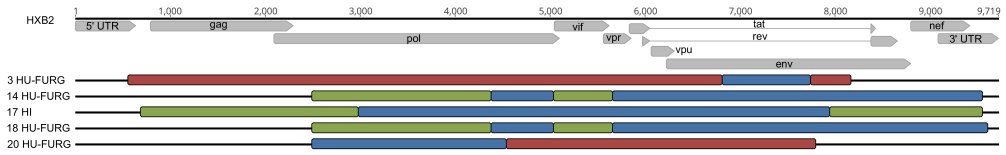


FIGURE 2 | Classification of HIV-1 recombinant viruses. The recombinant patterns were defined by phylogeny and similarity analyses. Each color represents a different subtype: red for subtype B, green for subtype F1, and blue for subtype C. Samples are identified before their respective virus structure and the HXB2 reference genome sequence is at the top of the Figure for reference positioning purpose.

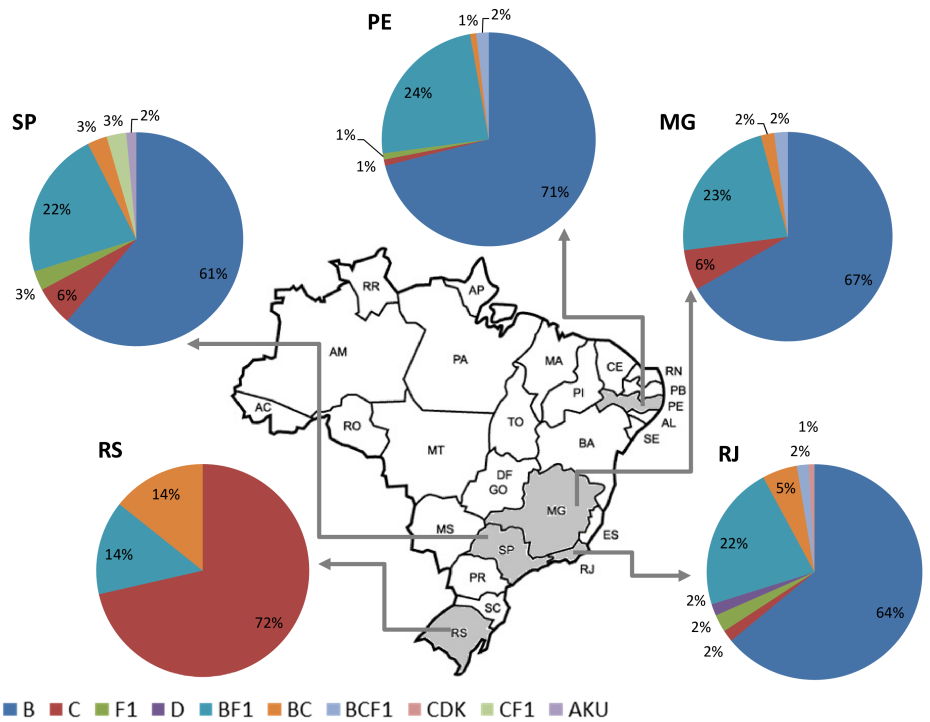


FIGURE 3 | Summary of the HIV-1 subtype distribution in the five Brazilian states represented in this pooled analysis (n = 350). Each pie graph represents one state as depicted in the map. HIV-1 subtypes and recombinants are color-coded according to the legend at the bottom of the figure.

unique URF_BF1 previously described (HI-28), and three were distinct URF_BC strains (one of those also described, HI-08) (**Figure 2**; Alves et al., 2017).

Altogether, the pooled analysis of all Brazilian NGS data, including NFLG (29%, 79/275) and partial sequences (29%, 22/75), showed a significant prevalence of recombinant forms (29%; 101/350). Considering the distribution of the HIV-1 subtypes and recombinant viruses in the five Brazilian states analyzed, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul), we could observe a highly diversified pattern of HIV-1 subtype distribution (**Figure 3**). A higher prevalence of recombinant forms in São Paulo and Rio de Janeiro (30%) could also be found, followed by Rio Grande do Sul (28%), and by Minas Gerais and Pernambuco (27%), although those differences were not significant (data not shown).

De novo analysis generated results suggestive of multiple infections (more than one IVA-contig in the same genomic region) for four samples (HU-UFRJ-03, two contigs; HI-11, two contigs; HI-14, three contigs, and HI-17, five contigs). For three of them (HU-UFRJ-03, HI-11, and HI-14), only one of the contigs generated intact open reading frames (ORFs), while the other contigs presented truncated ORFs showing multiple stop codons, consistent with APOBEC-mediated G-to-A mutations. They were confirmed as hypermutated sequences when compared to the viable sequence from the respective patient in Hypermut ($p < 0.05$, data not shown). Patient HI-17 had several overlapping regions between the contigs, two of them at the *gag-pol* region and three at *env* (**Figure 4A**). Phylogenetic trees comprising the contigs, the top-ten best hits found in BLASTn searches for each contig and reference sequences were constructed for each overlapping region. Overall, these trees showed different clustering profiles between the contigs and HIV-1 subtype references, suggesting the presence of variants with different HIV-1 subtypes within this samples (**Figure 4B**). The contigs were then submitted to Simplot

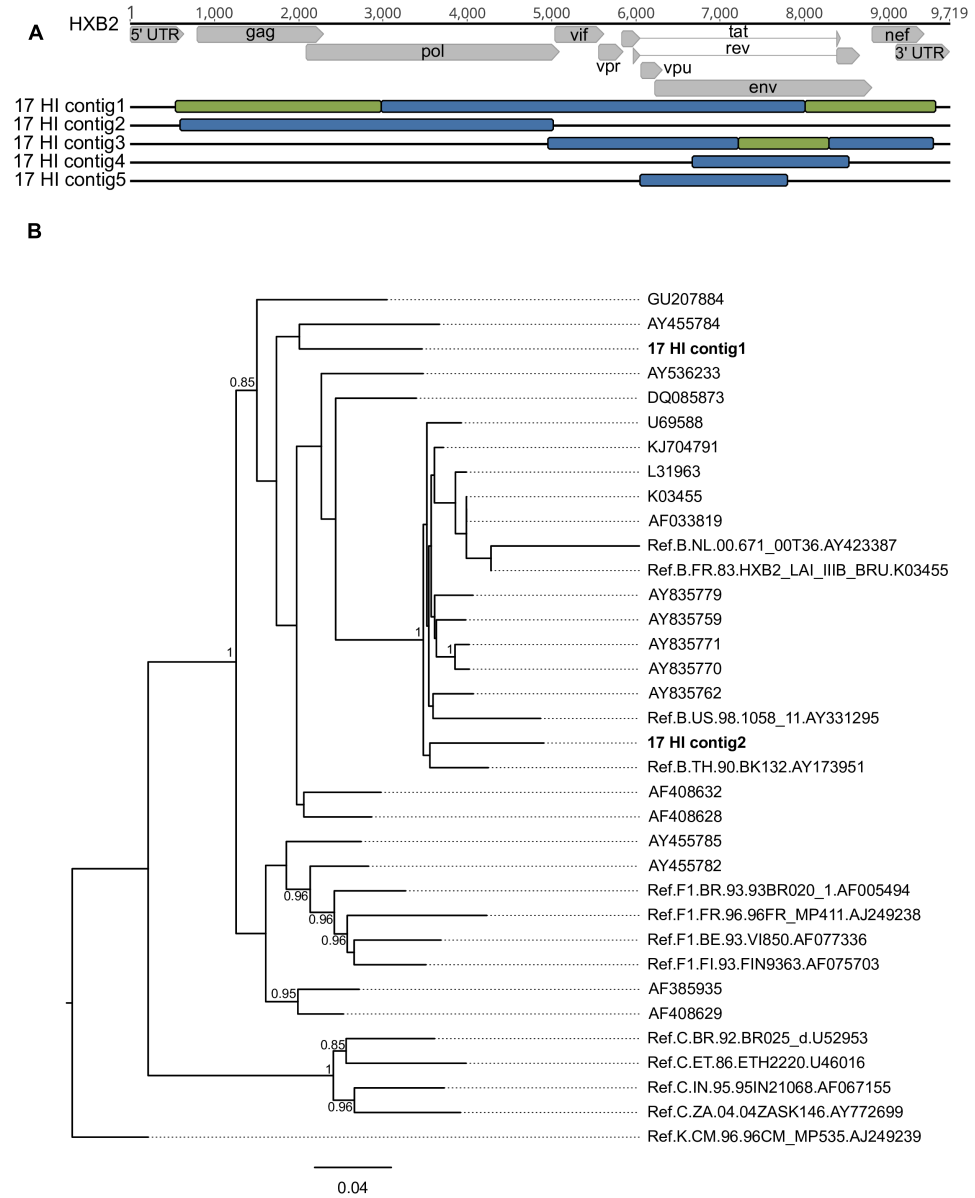


FIGURE 4 | Different contigs representing HIV-1 sequences found in patient HI-17 through multiple infection investigation with their respective positions along the HXB2 reference genome and phylogenetic classification **(A)** considering the bootscaning **(B)** and maximum likelihood phylogenetic analyses. Each color represents a different subtype: red for subtype B, green for subtype F1 and blue for subtype C. Samples are identified before their respective virus structure and the HXB2 reference genome sequence is at the top of the Figure for reference positioning purpose.

analysis to confirm the subtype classification and determine the recombination breakpoint profile. All contigs were confirmed as distinct variants. For the HI-17 patient, the longest contig had the same recombination profile observed for the consensus sequence (URF_BF1), one had a distinct recombination profile comprising B and F1 subtypes, and three were classified as subtype B (**Figure 4A**). Another two cases of multiple infections with distinct subtypes were described by Pessoa et al., one involving viruses of subclade F1 and subtype B and another involving a CBF1 recombinant and a non-recombinant subtype B (Pessoa et al., 2015, 2016).

DISCUSSION

The present study describes the HIV-1 genetic diversity and molecular epidemiology observed in Brazil using NGS-generated HIV-1 sequences, combining recently published reports and novel data from our group. In comparison to our previous published study with HIV$^+$ patients recruited at HFI, we extended our analysis for three of the nine patients which data could not be obtained in the first study, and obtained NFLG sequences for another five patients (Alves et al., 2017). Concerning the integrity of the open reading frames (ORF),

4.5% (11/247) of the sequences available in the literature, and 25% (7/28) of our sequences displayed mutations and/or insertions and deletions resulting in frameshifts or premature stop codons.

Considering our cohort, a high prevalence of infection by HIV-1 subtype B viruses was found among patients from Rio de Janeiro (79%, 27/34). On the other hand, subtype C was the most prevalent subtype in Rio Grande do Sul, with a similarly high proportion (72%, 10/14), highlighting the regional differences observed in HIV-1 subtype distribution in the country. It is well documented that the overall prevalence of non-B strains, such as URF_BF1, URF_BC, and particularly subtype C and CRF31_BC in the South of Brazil, has been increasing (Santos et al., 2007; Cardoso et al., 2009; Machado et al., 2009; Almeida et al., 2012; Velasco-de-Castro et al., 2014). Similarly to some recently published data using NFLG, the recombinants identified in our cohort did not show any similarity with the CRFs already described (Sanabani et al., 2013; Alves et al., 2017). However, it is worth to mention that two partial sequences from HU-FURG show the same recombinant pattern (14 HU-FURG and 18 HU-FURG) and did not have any evidence of epidemiological linkage between them.

Through the analysis of HIV-1-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul), we could determine, with unprecedented accuracy, the prevalence of HIV-1 recombinant forms in the Brazilian epidemic and their distribution across the country. The high prevalence of recombinant strains (29%) identified by NGS is supported by the circulation of multiple subtypes and consistent with the hypothesis that novel recombinants are continuously arising in the Brazilian epidemic (Sanabani et al., 2013). Our cohort presented 19% of recombinant sequences, a lower prevalence than 28 and 40% observed in other Brazilian NGS-based studies (Pessoa et al., 2015, 2016). The higher prevalence of recombinant forms when compared to recent Brazilian Sanger sequencing-based studies (ranging from 5 to 16%) can be attributed to the smaller genomic region analyzed in the latter, mostly based only on *pol* gene, which impairs the accurate classification of recombinants (Librelotto et al., 2015; Moura et al., 2015; da Costa et al., 2016; Dos Anjos Silva et al., 2016; Delatorre et al., 2017; Filho and Brites, 2017; Lima et al., 2017). Some of these studies covering small genomic regions could not find recombinant strains, like the study conducted by Corado et al. (2017) among 73 individuals from Roraima state, northern Brazil. A study conducted by Graf et al. (2016) found a higher proportion of recombinant strains than others Sanger-based studies (21%, 66/317). However, this study used molecular data from more than one gene (HIV-1 *pol*, *env*, or both). It should also be noted that 30% of the URF_BC samples were intergenic recombinants whose recombination breakpoints were not documented within these fragments (Graf et al., 2016). The comparison of the estimated prevalence of recombinant virus between the classical Sanger-based approach and the NGS data clearly highlights underestimated rates in the former analyses, mainly associated with the smaller genomic region analyzed, which also implies in the inaccurate detection of the recombinant breakpoints and, therefore, their classification.

HIV-1 NFLG-based studies can unveil an underestimated rate of recombinant viruses in the country. Using NFLG data, several studies have described new Brazilian CRF strains designated as CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF, CRF46_BF, and CRF31_BC, CRF70_BF1, CRF71_BF1, CRF72_BF1, CRF90_BF1 (De Sa Filho et al., 2006; Sanabani et al., 2006a,b, 2010; Santos et al., 2006; Guimaraes et al., 2008; Pessoa et al., 2014a,b; Reis et al., 2017). An important study conducted by Pessôa et al. evaluated the complete genomes of HIV-1 strains by NGS previously sequenced by Alencar et al. assigned to subtype F1 and showed that 23 of the 24 samples analyzed were BF recombinants, with 4 CRF70_BF1 and 11 CRF71_BF1 novel recombinant types (Alencar et al., 2013; Pessoa et al., 2014b). The same was observed by Marques et al. (2018) where 34 of the 55 sequences analyzed were classified as recombinants. In our cohort, only one sequence was classified as subtype F1 and three as URF-BF1, which corroborates to literature and highlights the higher prevalence of recombinants forms comprising subtype F1.

Regarding molecular diversity, our pooled analysis corroborates the crescent prevalence of non-B strains in the Brazilian epidemic, confirming the phylogenetic intermixing of HIV-1 sequences. The most prominent case comprises subtype C and C-containing recombinant forms expanding from the South of Brazil to other regions (Bello et al., 2012; Graf and Pinto, 2013; Graf et al., 2015). Non-B strains represent 39% of the sequences from São Paulo, 36% from Rio de Janeiro, 33% from Minas Gerais, and 29% from Pernambuco. Subtype B was not found in Rio Grande do Sul, probably because of the small number of samples analyzed in this region.

We also investigated infection by distinct variants using a *de novo* strategy to obtain sequences from each patient that are subsequently run in one of the reference-guided approaches using this sequence as a reference (see section Data Analysis of Materials and Methods). This strategy was employed by several studies to reduce the influence of a reference genome in the assembly process while investigating multiple infections (Mangul et al., 2014; Aralaguppe et al., 2016; Alampalli et al., 2017; Baaijens et al., 2017). The prevalence of multiple infections observed in our study (2%, 1/48) was similar to the prevalence reported by Pessôa et al. (4%, 1/24 and 2%, 1/47), but both were greater than the prevalence subsequently reported (0.3%, 1/259) (Pessoa et al., 2014b, 2015, 2016, respectively). At this point, it is not possible to infer whether the distinct viral strains resulted from coinfections or the acquisition of a second variant after the establishment of the first one (superinfections).

We are aware that the pooled analysis presented here includes HIV-positive patients with different HIV clinical profiles. While our cohort is composed by patients under first-line HAART and undetectable HIV viral load for at least 12 months prior to collection date attending at sexually transmitted diseases/HIV ambulatory, the studies conducted by Pessôa et al. involved recently infected donors at four blood centers. However, it should be noted that all epidemiological HIV-1 NFLG studies based on NGS conducted in Brazil available in the literature were

included in this pooled analysis. However, it should be noted that our convenience samples may have biased the analysis, especially regarding the multiple infection prevalence, which requires analysis from a larger data set.

Like previously described, the inclusion criteria used in this study were very strict and represented an important barrier to enroll a large number of patients. The difficulty at PCR-amplification of archived proviral genomes was also a limitation due to early chronic infection and undetectable HIV viral load. It is also important to mention that HU-FURG samples were incorporated in this study at a later stage. This fact, coupled with the difficulty in the PCR amplification of these samples resulted in a limited number of sequences from Rio Grande do Sul and additional studies are necessary to complement our findings. Another limitation of our study was that only the NFLG sequences ($n = 28$) were evaluated for ORF intactness, and seven of them (25%) had stop codons due to hypermutation or to frameshift deletions.

The analysis of all Brazilian HIV-1 NFLG obtained by NGS give us a more accurate evaluation of the viral diversity present in this epidemic. Through the subtype analyses conducted in this large cohort, we were able to find high rates of viral recombination, showing that larger viral genomic regions are required for reliable genetic evaluation and thus to establish effective public health policies to assure suitable HIV screening, diagnosis, monitoring and novel strategies based on viral variability. Our data highlight the great HIV genetic diversity found in Brazil and unveils a more accurate scenario of the HIV evolutionary dynamics in the region.

ETHICS STATEMENT

This study was carried out and approved in accordance with the recommendations of the Ethics Committees in Research of the Brazilian National Cancer Institute – INCA and of HFI (CAAE 52862016.9.0000.5274), HUCFF-UFRJ (CAAE 56604816.2.0000.5257), and HU-FURG (CAAE 52862016.9.3001.5324), with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

BA, ES, and MS conceived the study. BA and VDH collected and processed the samples. BA, JS, IP, and OB performed all molecular biology experiments and all the bioinformatics analyses. SS contributed with previously published HIV-1 NGS sequences from Brazil. PR-P and HF contributed reagents and provided critical reading of the manuscript. BA, JS, and MS wrote the manuscript. All authors read the manuscript and agreed with its final version and submission.

FUNDING

This work was supported by the Rio de Janeiro State Science Foundation – FAPERJ (Grant # E-26/201.331/2014 to MS), by the Brazilian National Research Council – CNPq (Grants # 454623/2014-3 to MS and # 460346/2014-8 to ES), and by MSD Avenir (Grant DS-2016-0005 to HF with a subcontract to MS). Intramural funding by the Brazilian Ministry of Health through INCA in the form of fellowships was provided to BA and JS at the post-doc level, to IP at the Ph.D. level and to OB at the undergraduate level.

ACKNOWLEDGMENTS

We would like to thank the medical staff personnel of Hospital Federal de Ipanema (Marianne M. Garrido and Sayonara R. Ribeiro), Hospital Universitário Dr. Miguel Riet Corrêa Jr. (Jussara Silveira and Rossana Basso), and Hospital Universitário Clementino Fraga Filho (Elizabeth S. Machado) for providing clinical follow-up and information on the studied patients at their respective clinical sites.

REFERENCES

Alampalli, S. V., Thomson, M. M., Sampathkumar, R., Sivaraman, K., Aj, U. K. J., Dhar, C., et al. (2017). Deep sequencing of near full-length HIV-1 genomes from plasma identifies circulating subtype C and infrequent occurrence of AC recombinant form in Southern India. *PLoS One* 12:e0188603. doi: 10.1371/journal.pone.0188603

Alencar, C. S., Sabino, E. C., Carvalho, S. M., Leao, S. C., Carneiro-Proietti, A. B., Capuani, L., et al. (2013). HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: role of infected blood donors as sentinel populations for molecular surveillance of HIV. *J. Acquir. Immune Defic. Syndr.* 63, 387–392. doi: 10.1097/QAI.0b013e31828ff979

Almeida, S. E., De Medeiros, R. M., Junqueira, D. M., Graf, T., Passaes, C. P., Bello, G., et al. (2012). Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. *Virol. J.* 9:306. doi: 10.1186/1743-422X-9-306

Alves, B. M., Siqueira, J. D., Garrido, M. M., Botelho, O. M., Prellwitz, I. M., Ribeiro, S. R., et al. (2017). Characterization of HIV-1 near full-length proviral genome quasispecies from patients with undetectable viral load undergoing first-line haart therapy. *Viruses* 9:E392.

Aralaguppe, S. G., Siddik, A. B., Manickam, A., Ambikan, A. T., Kumar, M. M., Fernandes, S. J., et al. (2016). Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. *J. Virol. Methods* 236, 98–104. doi: 10.1016/j.jviromet.2016.07.010

Baaijens, J. A., Aabidine, A. Z. E., Rivals, E., and Schonhuth, A. (2017). De novo assembly of viral quasispecies using overlap graphs. *Genome Res.* 27, 835–848. doi: 10.1101/gr.215038.116

Bello, G., Zanotto, P. M., Iamarino, A., Graf, T., Pinto, A. R., Couto-Fernandez, J. C., et al. (2012). Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. *PLoS One* 7:e35649. doi: 10.1371/journal.pone.0035649

Cardoso, L. P., Queiroz, B. B., and Stefani, M. M. (2009). HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naive patients from Central West Brazil. *J. Clin. Virol.* 46, 134–139. doi: 10.1016/j.jcv.2009.07.009

Corado, A. L., Bello, G., Leao, R. A., Granja, F., and Naveca, F. G. (2017). HIV-1 genetic diversity and antiretroviral drug resistance among individuals from

Roraima state, northern Brazil. *PLoS One* 12:e0173894. doi: 10.1371/journal.pone.0173894

da Costa, C. M., Costa De Oliveira, C. M., Chehuan, De Melo, Y. F., Delatorre, E., Bello, G., et al. (2016). High HIV-1 genetic diversity in patients from northern brazil. *AIDS Res. Hum. Retrovir.* 32, 918–922. doi: 10.1089/AID.2016.0044

de Medeiros, R. M., Junqueira, D. M., Matte, M. C., Barcellos, N. T., Chies, J. A., and Matos Almeida, S. E. (2011). Co-circulation HIV-1 subtypes B, C, and CRF31_BC in a drug-naive population from Southernmost Brazil: analysis of primary resistance mutations. *J. Med. Virol.* 83, 1682–1688. doi: 10.1002/jmv.22188

De Sa Filho, D. J., Sucupira, M. C., Caseiro, M. M., Sabino, E. C., Diaz, R. S., and Janini, L. M. (2006). Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res. Hum. Retroviruses* 22, 1–13.

Delatorre, E., Couto-Fernandez, J. C., and Bello, G. (2017). HIV-1 Genetic diversity in Northeastern Brazil: high prevalence of non-b subtypes. *AIDS Res. Hum. Retroviruses* 33, 639–647. doi: 10.1089/AID.2017.0045

Dos Anjos Silva, L., Divino, F., Da Silva Rego, M. O., Lima Lopes, I. G., Nobrega Costa, C. M., et al. (2016). HIV-1 genetic diversity and transmitted drug resistance in antiretroviral treatment-naive individuals from Amapa State, Northern Brazil. *AIDS Res. Hum. Retroviruses* 32, 373–376. doi: 10.1089/AID.2015.0280

Dudley, D. M., Bailey, A. L., Mehta, S. H., Hughes, A. L., Kirk, G. D., Westergaard, R. P., et al. (2014). Cross-clade simultaneous HIV drug resistance genotyping for reverse transcriptase, protease, and integrase inhibitor mutations by Illumina MiSeq. *Retrovirology* 11:122. doi: 10.1186/s12977-014-0122-8

Filho, A. W. O., and Brites, C. (2017). Geolocalization of HIV-1 subtypes and resistance mutations of patients failing antiretroviral therapy in Salvador - Brazil. *Braz. J. Infect. Dis.* 21, 234–239. doi: 10.1016/j.bjid.2017.02.006

Graf, T., Machado Fritsch, H., De Medeiros, R. M., Maletich Junqueira, D., Esteves De Matos Almeida, S., and Pinto, A. R. (2016). Comprehensive characterization of HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS. *J. Virol.* 90, 8160–8168. doi: 10.1128/JVI.00363-16

Graf, T., and Pinto, A. R. (2013). The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 435, 170–178. doi: 10.1016/j.virol.2012.08.048

Graf, T., Vrancken, B., Maletich Junqueira, D., De Medeiros, R. M., Suchard, M. A., Lemey, P., et al. (2015). Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *J. Virol.* 89, 12341–12348. doi: 10.1128/JVI.01681-15

Guimaraes, M. L., Eyer-Silva, W. A., Couto-Fernandez, J. C., and Morgado, M. G. (2008). Identification of two new CRF_BF in rio de Janeiro State, Brazil. *AIDS* 22, 433–435.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and 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. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Hemelaar, J. (2013). Implications of HIV diversity for the HIV-1 pandemic. *J. Infect.* 66, 391–400. doi: 10.1016/j.jinf.2012.10.026

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and Who-Unaids Network for Hiv Isolation, and Characterisation. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 25, 679–689. doi: 10.1097/QAD.0b013e328342ff93

Hunt, M., Gall, A., Ong, S. H., Brener, J., Ferns, B., Goulder, P., et al. (2015). IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* 31, 2374–2376. doi: 10.1093/bioinformatics/btv120

Junqueira, D. M., and Almeida, S. E. (2016). HIV-1 subtype B: Traces of a pandemic. *Virology* 495, 173–184. doi: 10.1016/j.virol.2016.05.003

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., and Mclnerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6:29.

Librelotto, C. S., Graf, T., Simon, D., De Almeida, S. E., and Lunge, V. R. (2015). HIV-1 epidemiology and circulating subtypes in the countryside of South Brazil. *Rev. Soc. Bras. Med. Trop.* 48, 249–257. doi: 10.1590/0037-8682-0083-2015

Lima, K., Leal, E., Cavalcanti, A. M. S., Salustiano, D. M., De Medeiros, L. B., Da Silva, S. P., et al. (2017). Increase in human immunodeficiency virus 1 diversity and detection of various subtypes and recombinants in north-eastern Brazil. *J. Med. Microbiol.* 66, 526–535. doi: 10.1099/jmm.0.000447

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., and Novak, N. G. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152–160.

Machado, L. F., Costa, I. B., Folha, M. N., Da Luz, A. L., Vallinoto, A. C., Ishak, R., et al. (2017). Lower genetic variability of HIV-1 and antiretroviral drug resistance in pregnant women from the state of Para, Brazil. *BMC Infect. Dis.* 17:270. doi: 10.1186/s12879-017-2392-y

Machado, L. F., Ishak, M. O., Vallinoto, A. C., Lemos, J. A., Azevedo, V. N., Moreira, M. R., et al. (2009). Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the amazon region of Brazil. *AIDS Res. Hum. Retroviruses* 25, 961–966. doi: 10.1089/aid.2009.0027

Mangul, S., Wu, N. C., Mancuso, N., Zelikovsky, A., Sun, R., and Eskin, E. (2014). Accurate viral population assembly from ultra-deep sequencing data. *Bioinformatics* 30, i329–i337. doi: 10.1093/bioinformatics/btu295

Marques, B. C. L., Morgado, M. G., and Guimaraes, M. L. (2018). Potential overestimation of HIV-1 sub-subtype F1 circulation in Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* 113:e170483. doi: 10.1590/0074-02760170483

Moura, M. E., Da Guarda Reis, M. N., Lima, Y. A., Eulalio, K. D., Cardoso, L. P., and Stefani, M. M. (2015). HIV-1 transmitted drug resistance and genetic diversity among patients from Piaui State, Northeast Brazil. *J. Med. Virol.* 87, 798–806. doi: 10.1002/jmv.24087

Ode, H., Matsuda, M., Matsuoka, K., Hachiya, A., Hattori, J., Kito, Y., et al. (2015). Quasispecies analyses of the HIV-1 near-full-length genome with illumina MiSeq. *Front. Microbiol.* 6:1258. doi: 10.3389/fmicb.2015.01258

Overbaugh, J., and Bangham, C. R. (2001). Selection forces and constraints on retroviral sequence variation. *Science* 292, 1106–1109.

Pessoa, R., Carneiro Proietti, A. B., Busch, M. P., and Sanabani, S. S. (2014a). Identification of a Novel HIV-1 circulating recombinant form (CRF72_BF1) in deep sequencing data from blood donors in Southeastern Brazil. *Genome Announc.* 2, e386–e314. doi: 10.1128/genomeA.00386-14

Pessoa, R., Watanabe, J. T., Calabria, P., Felix, A. C., Loureiro, P., Sabino, E. C., et al. (2014b). Deep sequencing of HIV-1 near full-length proviral genomes identifies high rates of BF1 recombinants including two novel circulating recombinant forms (CRF) 70_BF1 and a disseminating 71_BF1 among blood donors in Pernambuco, Brazil. *PLoS One* 9:e112674. doi: 10.1371/journal.pone.0112674

Pessoa, R., Loureiro, P., Esther Lopes, M., Carneiro-Proietti, A. B., Sabino, E. C., Busch, M. P., et al. (2016). Ultra-deep sequencing of HIV-1 near full-length and partial proviral genomes reveals high genetic diversity among brazilian blood donors. *PLoS One* 11:e0152499. doi: 10.1371/journal.pone.0152499

Pessoa, R., Watanabe, J. T., Calabria, P., Alencar, C. S., Loureiro, P., Lopes, M. E., et al. (2015). Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. *Transfusion* 55, 980–990. doi: 10.1111/trf.12936

Reis, M., Bello, G., Guimaraes, M. L., and Stefani, M. M. A. (2017). Characterization of HIV-1 CRF90_BF1 and putative novel CRFs_BF1 in Central West, North and Northeast Brazilian regions. *PLoS One* 12:e0178578. doi: 10.1371/journal.pone.0178578

Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* 242, 1171–1173.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–56.

Sanabani, S., Kleine Neto, W., Kalmar, E. M., Diaz, R. S., Janini, L. M., and Sabino, E. C. (2006a). Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. *Infect. Genet. Evol.* 6, 368–377.

Sanabani, S., Neto, W. K., De Sa Filho, D. J., Diaz, R. S., Munerato, P., Janini, L. M., et al. (2006b). Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res. Hum. Retroviruses* 22, 171–176.

Sanabani, S. S., Pastena, E. R., Neto, W. K., Martinez, V. P., and Sabino, E. C. (2010). Characterization and frequency of a newly identified HIV-1 BF1 intersubtype

circulating recombinant form in Sao Paulo, Brazil. *Virol. J.* 7:74. doi: 10.1186/1743-422X-7-74

Sanabani, S. S., Pessoa, R., Soares De Oliveira, A. C., Martinez, V. P., Giret, M. T., De Menezes Succi, R. C., et al. (2013). Variability of HIV-1 genomes among children and adolescents from Sao Paulo, Brazil. *PLoS One* 8:e62552. doi: 10.1371/journal.pone.0062552

Santoro, M. M., and Perno, C. F. (2013). HIV-1 genetic variability and clinical implications. *ISRN Microbiol.* 2013:481314. doi: 10.1155/2013/481314

Santos, A. F., Schrago, C. G., Martinez, A. M., Mendoza-Sassi, R., Silveira, J., Sousa, T. M., et al. (2007). Epidemiologic and evolutionary trends of HIV-1 CRF31_BC-related strains in southern Brazil. *J. Acquir. Immune Defic. Syndr.* 45, 328–333.

Santos, A. F., Sousa, T. M., Soares, E. A., Sanabani, S., Martinez, A. M., Sprinz, E., et al. (2006). Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* 20, 2011–2019.

Thomson, M. M., and Najera, R. (2005). Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 7, 210–224.

UNAIDS (2018). *The Joint United Nations Programme on HIV/AIDS (UNAIDS). Data 2018*. Geneva: UNAIDS.

Velasco-de-Castro, C. A., Grinsztejn, B., Veloso, V. G., Bastos, F. I., Pilotto, J. H., Fernandes, N., et al. (2014). HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. *PLoS One* 9:e87622. doi: 10.1371/journal.pone.0087622

Yerly, S., Jost, S., Monnat, M., Telenti, A., Cavassini, M., Chave, J. P., et al. (2004). HIV-1 co/super-infection in intravenous drug users. *AIDS* 18, 1413–1421.

Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., et al. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* 76, 11273–11282.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 April 2019
doi: 10.3389/fmicb.2019.00769

Local Epidemics Gone Viral: Evolution and Diffusion of the Italian HIV-1 Recombinant Form CRF60_BC

Alessia Lai[1], Francesco Roberto Simonetti[2], Gaetano Brindicci[3], Annalisa Bergna[1], Simona Di Giambenedetto[4], Gaetana Sterrantino[5], Cristina Mussini[6], Stefano Menzo[7], Patrizia Bagnarelli[7], Maurizio Zazzi[8], Gioacchino Angarano[3], Massimo Galli[1], Laura Monno[3] and Claudia Balotta[1]*

[1] Department of Biomedical and Clinical Sciences L. Sacco, University of Milan, Milan, Italy, [2] Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [3] Clinic of Infectious Diseases, University of Bari Aldo Moro, Bari, Italy, [4] Institute of Infectious Diseases, Catholic University of Sacred Heart, Rome, Italy, [5] Division of Tropical and Infectious Diseases, Careggi Hospital, Florence, Italy, [6] Clinic of Infectious Diseases, University of Modena and Reggio Emilia, Modena, Italy, [7] Unit of Virology, Azienda Ospedaliero-Universitaria 'Ospedali Riuniti', Torrette, Italy, [8] Department of Medical Biotechnologies, University of Siena, Siena, Italy

OPEN ACCESS

Edited by:
Joris Hemelaar, University of Oxford, United Kingdom

Reviewed by:
Dimitrios Paraskevis, National and Kapodistrian University of Athens, Greece
Antoinette Van Der Kuyl, University of Amsterdam, Netherlands
Marta Canuti, Memorial University of Newfoundland, Canada

***Correspondence:**
Alessia Lai
alessia.lai@unimi.it

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 October 2018*
Accepted: *26 March 2019*
Published: *12 April 2019*

Citation:
Lai A, Simonetti FR, Brindicci G, Bergna A, Di Giambenedetto S, Sterrantino G, Mussini C, Menzo S, Bagnarelli P, Zazzi M, Angarano G, Galli M, Monno L and Balotta C (2019) Local Epidemics Gone Viral: Evolution and Diffusion of the Italian HIV-1 Recombinant Form CRF60_BC. Front. Microbiol. 10:769. doi: 10.3389/fmicb.2019.00769

The molecular epidemiology of HIV-1 in Italy is becoming increasingly complex, mainly due to the spread of non-B subtypes and the emergence of new recombinant forms. We previously characterized the outbreak of the first Italian circulating recombinant form (CRF60_BC), occurring among young MSM living in Apulia between the years 2009 and 2011. Here we show a 5-year follow-up surveillance to trace the evolution of CRF60_BC and to investigate its further spread in Italy. We collected additional sequences and clinical data from patients harboring CRF60_BC, enrolled at the Infectious Diseases Clinic of the University of Bari. In addition to the 24 previously identified sequences, we retrieved 27 CRF60_BC sequences from patients residing in Apulia, whose epidemiological and clinical features did not differ from those of the initial outbreak, i.e., the Italian origin, young age at HIV diagnosis (median: 24 years; range: 18–37), MSM risk factor (23/25, 92%) and recent infection (from 2008 to 2017). Sequence analysis revealed a growing overall nucleotide diversity, with few nucleotide changes that were fixed over time. Twenty-seven additional sequences were detected across Italy, spanning multiple distant regions. Using a BLAST search, we also identified a CRF60_BC sequence isolated in United Kingdom in 2013. Three patients harbored a unique second generation recombinant form in which CRF60_BC was one of the parental strains. Our data show that CRF60_BC gained epidemic importance, spreading among young MSM in multiple Italian regions and increasing its population size in few years, as the number of sequences identified so far has triplicated since our first report. The observed further divergence of CRF60_BC is likely due to evolutionary bottlenecks and host adaptation during transmission chains. Of note, we detected three second-generation recombinants, further supporting a widespread circulation of CRF60_BC and the increasing complexity of the HIV-1 epidemic in Italy.

Keywords: HIV-1 recombinant forms, HIV-1 molecular epidemiology, HIV-1 evolution, HIV-1 outbreak, second generation recombinants

INTRODUCTION

The human immunodeficiency virus (HIV-1) is a well-known moving target. HIV-1 high variability and rapid adaptation impose an additional challenge to its diagnosis, treatment, and prevention (Hemelaar, 2013). However, HIV-1 evolutionary potential can be harnessed to understand its complex dynamics, both at cellular and population level. Thanks to viral diversity, molecular epidemiology allows to track HIV-1 global and local epidemics and provide insights on how to control their further diffusion (Hemelaar, 2012).

The pandemic group M spread worldwide and differentiated over time into nine distinct subtypes (A–D, F–H, J, and K) and 8 sub-subtypes (A1–A6, F1, and F2) (Peeters et al., 2014). In addition to HIV-1 error prone reverse transcriptase and high replicative potential, recombination plays a key role in its variability (Hu and Hughes, 2012). Due to recombination, when two or more subtypes co-circulate among high risk groups in the same geographical area, unique recombinant forms (URFs) emerge and, if they are detected in at least three epidemiologically unrelated patients, they gain the role of circulating recombinant forms (CRFs) (Robertson et al., 1995). The number of identified CRFs has grown dramatically; currently, 98 CRFs have been characterized and their virological and epidemiological data are available at the Los Alamos HIV database[1]. The first CRFs have been found mainly in central Africa and South America, where multiple subtypes co-circulate at high prevalence (Thomson et al., 2002). In the past few years, some of these recombinant progenies have caused outbreaks (Monno et al., 2012; Gonzalez-Domenech et al., 2018) and in some cases have shown an increased fitness (Rawson et al., 2018). For example, the Cuban CRF19_cpx, has been associated with a higher viral load, a more rapid progression to AIDS (Kouri et al., 2015) and antiretroviral resistance mutations (Kouri et al., 2014). HIV-1 M chimeric forms, URFs and CRFs altogether, are estimated to represent at least 20% of infections globally, with CRF01_AE and CRF02_AG being the most prevalent, accounting for about 13% of all infections (Osmanov et al., 2002). Despite their origin, some CRFs gained epidemiological relevance in previously B subtype-restricted countries with lower HIV-1 prevalence. Within the last years, multiple new CRFs have been identified in European countries such as Russia (Liitsola et al., 1998; Lukashov et al., 1999; Baryshev et al., 2014), Portugal, Spain (Delgado et al., 2002; Fernandez-Garcia et al., 2016), Italy (Simonetti et al., 2014), United Kingdom (Foster et al., 2014), Luxembourg (Struck et al., 2015), and France (Leoz et al., 2013; Recordon-Pinson et al., 2018).

The molecular epidemiology of HIV-1 in Italy, previously a B-restricted area, has become complex, as migratory waves and travels favored the penetration of HIV-1 forms from countries where non-B subtypes prevail (Lai et al., 2012, 2014a,b, 2016). A comprehensive study reported that non-B variants increased dramatically since the beginning of the epidemic with a prevalence of 15.8 and 29.7% of all HIV-1 diagnoses from 2006 to 2016 (Rossetti et al., 2018). After subtype B, the three most abundant pure subtypes in Italy were F, A and C, and CRFs represented almost 50% of non-B forms (Lai et al., 2010). Monno et al. (2012) reported a sudden HIV-1 outbreak among young men-who-have-sex-with-men (MSM) residing in Apulia, a region of southern Italy. This outbreak was caused by an inter-subtype recombinant, which was identified as the first Italian CRF, named CRF60_BC, by a subsequent study (Simonetti et al., 2014). This recombinant form is mostly composed of subtype C with the exception of three fragments from a parental B subtype which covered a part of the 5′ LTR (HXB2 positions 0–462), a small region of RT *pol* gene (HXB2 positions 2,804–3,037) and the last part of *env* gene. *Nef* gene and most of the 3′ LTR (HXB2 positions 8,662–9,548) were also subtype B. At the time of its characterization, the outbreak involved 22 individuals residing in Apulia and additional few patients from Emilia-Romagna, Tuscany and Sicily, all diagnosed between 2009 and 2011.

Here, we present a follow up study in which we investigated the further circulation and evolution of CRF60_BC. We combined epidemiological data with phylogenetic analyses on sequences retrieved in Apulia and those available from a nation-wide database. We observed an increasing population size and evolution of CRF60_BC, whose spread resulted also in novel recombination events. Overall, the present study portrays an increasingly complex and dynamic landscape of the HIV-1 Italian epidemic.

MATERIALS AND METHODS

Patients

We collected sequences and clinical data from patients, harboring CRF60_BC, enrolled at the Infectious Diseases Clinic of the University of Bari. We found additional sequences from the national ARCA cohort. Putative CRF60_BC sequences were searched among those assigned to subtype C or CRF31_BC in *pol* gene, according to a preliminary subtyping based on BLAST. As the CRF60_BC contains a small portion of B subtype in this region, the BLAST search lead to a subtype misclassification. Sequences were retrieved from the genotypic drug resistance assay performed by the local Services at diagnosis or prior to the start of therapy or at treatment failure. Only the first available resistance genotype was considered for downstream analyses. Genotypic resistance from plasma HIV-RNA was determined by bulk Sanger sequencing using commercially available or homebrew methods, depending on the contributing laboratory.

ARCA is an observational HIV cohort approved by the Regional Ethical Committee of Tuscany (Comitato Etico Area Vasta Toscana Sudest). The study was conducted in accordance with the 1964 Declaration of Helsinki and the ethical standards of the Italian Ministry of Health. The national ARCA database, currently contained demographic and viroimmunological data of 40,654 patients enrolled at 90 Clinical Centers, and at list one sequence before the start of antiretroviral therapy for all subjects. All patients belonging to the ARCA database provided informed consent to have their anonymized data stored on a central server and used for research studies. For all patients, epidemiological data (gender, risk category, country of origin, date of diagnosis,

[1]https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html

and age) were collected by physicians from medical records and then included in the databases together with virological, immunological, and treatment information through standard procedures cleared by the Southeast Tuscany Ethics Committee. When available, analyses were conducted on multiple HIV-1 regions: protease (PRO), reverse transcriptase (RT), integrase (INT), gp120 and gp41. The estimated time from infection was calculated by computing the fraction of ambiguous nucleotides in the PRO-RT sequences for all patients, as previously reported (Kouyos et al., 2011).

Alignment and Phylogenetic Analysis

BioEdit version 7.2.6.1[2] was used to manually align query sequences with pure subtypes and CRF reference sequences. The alignment of a comprehensive list of reference sequences of different subtypes (A1, A2, B, C, D, F1, F2, G, H, J, K) was retrieved from the Los Alamos HIV database[3]. In particular, we included 47, 37, 36, and 37 reference sequences for PRO/RT, INT, gp120, and gp41 fragments, respectively. The evolutionary model that best fitted the data was selected using the information criterion implemented in JmodelTest v2.1.7 (Posada, 2008). We then conducted phylogenetic analyses for subtyping on more than 10,000 sequences from the national ARCA cohort with maximum likelihood approach implemented in MEGA v7.0 program to retrieve additional CRF60_BC circulating in Italy.

The monophyletic nature of the sequences was evaluated by means of the MrBayes program (Huelsenbeck and Ronquist, 2001) using a general-time reversible (GTR) or Hasegawa-Kishino-Yano (HKY) model of nucleotide substitution, a proportion of invariant sites, and gamma distributed rates among sites for PRO, RT, gp120 and gp41 regions, respectively. For INT region HKY plus a proportion of invariant sites was used. The final alignment encompassed 1,259, 1,029, 514, and 858 nucleotides for PRO/RT, gp120, gp41, and INT regions, respectively.

A Markov Chain Monte Carlo (MCMC) search was made for 2×10^6 generations using tree sampling every 100^{th} generation and a highly conservative burn-in fraction of 50% to obtain a better signal of convergence. The analysis was run until reaching the average standard deviation < 0.01. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and a posterior consensus tree was generated after a 50% burn-in.

Pairwise distances were measured with MEGA v7.0 (Kumar et al., 2016), as the p-distance of each taxa from the CRF60_BC sequence harbored by the patient with the oldest time of diagnosis (patient BAV275, year 2005), with variance estimation performed using 1,000 bootstrap replicates. Spearman test of correlation between genetic distance and time of diagnosis was conducted with GraphPad Prism v7.0. Nucleotide differences across CFR60_BC sequences from Apulia were plotted with Highlighter (Keele et al., 2008)[4], with sequences sorted according to tree topology; ambiguous IUPAC codes were excluded from the graph.

Similarity and bootscanning plots implemented in Simplot software v3.5.2[5] were used to identify the subtypes involved in the recombination and their breakpoints to reconstruct recombination pattern. Maximum Likelihood phylogenetic trees were constructed for individual segments identified within each breakpoint with bootstrapping on 1,000 replicates using MEGA v7.0 program. Recombination was also characterized by split decomposition method and neighbor-net network generated with SplitsTree program[6]. Final trees were visualized and annotated with FigTree v1.4.0[7].

HIV BLAST tool[8] was used to retrieve sequences showing high similarity and the same BC mosaic pattern within the Los Alamos HIV Sequence Database. The prediction of HIV-1 co-receptor usage was performed using Geno2Pheno[9].

Phylodynamic Analysis

Dated trees of the CRF60_BC portions (INT, gp120, and gp41) not containing recombination breakpoints were estimated by BEAST software v1.8.4 (Drummond and Rambaut, 2007) using the previously selected models. For this purpose, only sequences with known date of collection were used. Chains were run for 10–30 million generations with sampling every 1,000–3,000 generations. The results were visualized in Tracer v1.5 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 v1.8.4 available in the BEAST software package. Final trees were visualized and edited with FigTree v1.4.0.

RESULTS

Evolution of the CRF60_BC Outbreak in Apulia

We first investigated whether the CRF60_BC had spread further locally, where the original outbreak occurred in 2009–2011. In addition to the 24 sequences previously described (Monno et al., 2012), we retrieved from local clinical centers, 27 new CRF60_BC sequences from patients residing in Apulia. The epidemiological and clinical features of the patients carrying this variant did not differ from those of the initial outbreak (**Table 1**). Similarly to the patients previously described, these individuals were Italian (100%) with young age at HIV-1 diagnosis (median: 26 years; range: 18–53) and all but one were MSM (25/26, 96%). The majority of these patients had a recent diagnosis in 2012–2017 and a CD4+ T-cell count at diagnosis above

[2]http://www.mbio.ncsu.edu/bioedit/bioedit.html

[3]www.hiv.lanl.gov

[4]http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html

[5]http://sray.med.som.jhmi.edu/SCRoftware/simplot/

[6]http://www.splitstree.org/

[7]http://tree.bio.ed.ac.uk/software/figtree/

[8]http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html

[9]https://coreceptor.geno2pheno.org/

TABLE 1 | Characteristics of patients harboring CRF60_BC.

Characteristics	CRF60_BC (Monno et al., 2012) (*n* = 24)	CRF60_BC new from Apulia (*n* = 27)	CRF60_BC new outside Apulia (*n* = 27)
Region [%; (n)]			
Apulia	100 (24)	100 (27)	–
Emilia	–	–	22.2 (6)
Tuscany	–	–	22.2 (6)
Sicilia	–	–	3.7 (1)
Marche	–	–	3.7 (1)
Liguria	–	–	3.7 (1)
Lombardy	–	–	3.7 (1)
Lazio	–	–	29.6 (8)
Trentino	–	–	3.7 (1)
Piemonte	–	–	3.7 (1)
Umbria	–	–	3.7 (1)
Unknown	–	–	3.7 (1)
Gender [%; (n)]			
Male	91.7 (22)	100 (27)	81.4 (22)
Unknown	8.3 (2)	–	14.8 (4)
Age at diagnosis [years]			
Median [IQR] (n)	25 [25–25](24)	26 [22.5–29](27)	25 [25–34.5](12)
Year of diagnosis			
Median [IQR] (n)	2010 [2010–2011](24)	2013 [2012–2015](27)	2013.5 [2013–2014](11)
Country of origin [%; (n)]			
Italy	100 (24)	100 (27)	59.2 (16)
Cote d'Avoire	–	–	3.7 (1)
United Kingdom	–	–	3.7 (1)
Unknown	–	–	33.3 (9)
Modality of transmission			
Men-who-have-sex-with-men	83.3 (20)	92.6 (25)	40.7 (11)
Other	16.7 (4)	3.7 (1)	3.7 (1)
Unknown	–	3.7 (1)	55.5 (15)
CD4 count*[cell/μl]			
Median [IQR] (n)	465.5 [393.3–659.5](24)	506 [420–672.5](27)	570 [467–637](13)
pVL at sequencing [copies/mL]			
Median [IQR] (n)	7,700 [3,150–32,500](24)	8,000 [4,700–26,000](27)	24,197 [7,528–56,582.5](12)
Time from seroconversion [%; (n)]			
Seroconversion within 12 month	54.5 (6)	60 (6)	66.7 (2)
Chronic	45.5 (5)	40 (4)	33.3 (1)
Estimated age of infection[%; (n)]**			
<1 year	75 (18)	77.8 (21)	40.7 (11)
>1 year	25 (6)	22.2 (6)	59.3 (16)

**At the time of sequencing; ***As in Kouyos et al. (2011).*

350 cells/μl in 23 out 27 cases (median: 506 cells/μl; range: 46–1,187). Three subjects (BAV718, BAV731, BAV851) had levels of plasma HIV-RNA above 10^6 copies/ml, suggesting that they were diagnosed before reaching a viral set point. These data support that most of the patients carrying the CRF60_BC variant were infected recently, during the last few years after the initial outbreak. However, we identified three subjects, BAV275, BAV334, BAV420, whose HIV-1 infection occurred before the initial outbreak, as their diagnosis dated back to 2005, 2006, and 2008, respectively (Simonetti et al., 2014). These patients were recorded in the ARCA database after the characterization of CRF60_BC. Last negative test was available for 10 subjects and confirmed a seroconversion time within 12 months in six cases. The analysis of the fraction of ambiguous nucleotides provided strong evidence of a recent infection (<1 year) for 77.8% ($n = 21$) of subjects. The Bayesian tree of patients from Apulia is shown

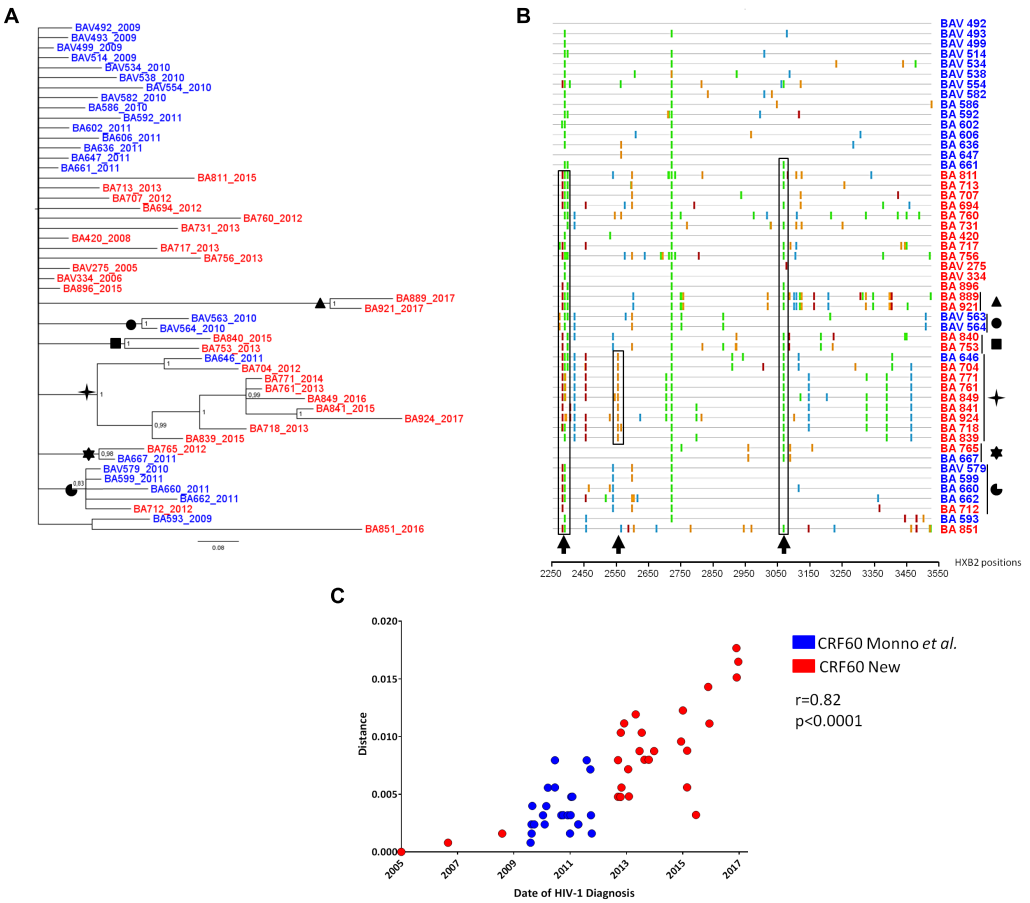


FIGURE 1 | Evolution of CRF60_BC PRO-RT sequences from Apulia. **(A)** Bayesian tree of PRO-RT Apulian sequences obtained by MrBayes program. Names of sequences include the year of sampling. The newly identified sequences are shown in red, while those previously described are shown in blue. Posterior probabilities > 0.8 are shown at nodes. **(B)** Highlighter plot of nucleotide mismatches in the protease and RT regions; sequences are sorted based on the topology of the tree in **(A)**, with the top sequence used as the master. Arrows and boxes indicate non-synonymous changes occurring in known epitopes under cytotoxic T lymphocyte (CTL) pressure. Symbols indicated the sequence clusters present in the tree. **(C)** Correlation plot of genetic pairwise distances of HIV-1 diagnosis over time.

in **Figure 1A**. Twelve out 27 (44.4%) recently identified strains showed high similarity with the majority of previous described sequences (15/24, 62.5%). We identified six significant clusters supported by a posterior probability higher than 0.80. Two of them corresponded to known partners or epidemiological links (pp = 1, BA889-BA921, BAV563-BAV564). The largest cluster (pp = 1) involved one of the sequences identified in 2011 and eight newly detected strains (year of diagnosis ranging between 2012 and 2017). Another large cluster (pp = 0.83) included four previously described CRF60_BC sequences and one sequence identified in 2012. The remaining two clusters (pp = 1 and pp = 0.98, respectively) involved two strains each; the first included two new sequences (BA840 and BA753) and the second a previously described strain (BA667) with a new one (BA765).

The emergence of new cases of HIV-1 infection due to CRF60_BC could be explained both by an increase of diagnoses after the initial outbreak (more effective sampling) and actual onward dissemination. To further understand whether CRF60_BC has been involved in ongoing transmission chains, we looked at sequence diversity and accumulation of mutations over time (**Figures 1B,C**). We observed the emergence of non-synonymous mutations strongly associated with CD8+ T-cell immune pressure against epitopes in *gag* (HXB2 position 498) and *pol* (HXB2 positions 69, 72, 128, and 314) (Addo et al., 2003; Allen et al., 2005). Some mutations were fixed over time, like in the case of the large transmission cluster encompassing sequences from 2011 to 2017, in which we observed reversion to subtype B consensus of a known conserved *pol* epitope (HXB2 *pol* 314) associated with HLA-I responses (Addo et al., 2003; Apps et al., 2013). In addition, when we estimated pairwise distances of all sequences ($n = 47$) from the oldest isolate (2005), we detected a growing overall diversity that strongly correlated with time of HIV-1 diagnosis ($r = 0.82$, $p < 0.0001$), further supporting ongoing evolution of CRF60_BC over time.

Diffusion of CRF60_BC Outside Apulia

To investigate the further diffusion of CRF60_BC, we conducted subtyping analyses on the ARCA database retrieving 25

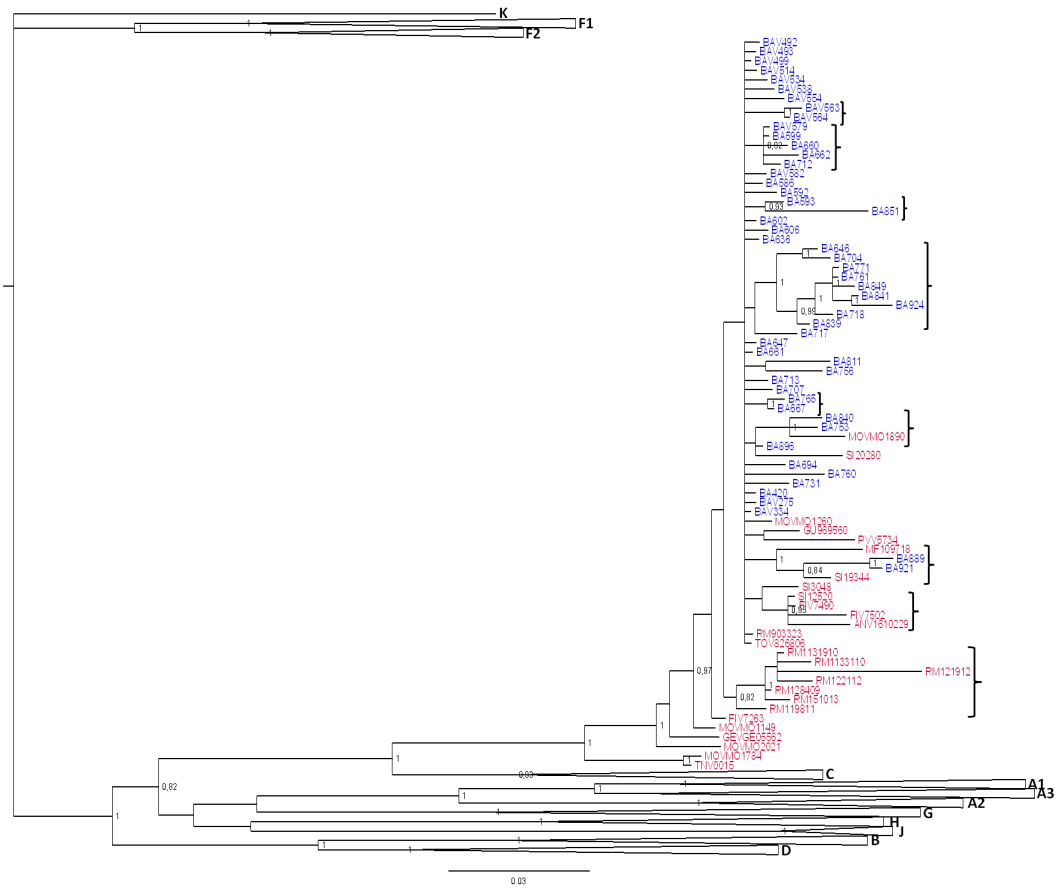


FIGURE 2 | Phylogenetic tree of CRF60_BC PRO-RT sequences detected in and outside of Apulia. Bayesian tree of PRO-RT CRF60_BC obtained using MrBayes program. Posterior probabilities > 0.8 are shown at nodes. Sequences in blue are from Apulia region, in red those retrieved from other Italian regions and two sequences detected using BLAST search. The brackets highlight the significant sub-clades.

sequences belonging to CRF60_BC (data not shown). **Figure 2** shows the Bayesian tree containing 78 sequences: the original and the newly identified sequences from Apulia (total, n = 51), sequences from different Italian regions obtained from the ARCA database (n = 25) and two sequences detected using BLAST search (GU969560 and MF109718).

Sequences assigned to CRF60_BC formed a highly supported cluster (pp = 1). Twenty-five sequences were detected across the Italian territory (Rome, Florence, Modena, Arezzo, Fermo, Pavia, Genoa, Perugia, Trento, and Turin). Some sub clusters were observed, containing 2 to 9 strains. Sequences from Apulia segregated alone, except for four sequences. Two strains (BA840 and BA753) grouped significantly (pp = 1) with a sequence from Modena (MOV-MO-1890) while the remaining two (BA889 and BA921) formed a sub cluster with a strain from Perugia (SI-19344) and a sequence detected from a patient in United Kingdom (MF109718). This latter sequence, retrieved by BLAST, had no demographic information apart from the sampling date (November 2013), preventing us to rule out whether the patient acquired the infection in his/her country of origin or abroad. All sequences from Rome (n = 7), with the exception of RM9-03323, formed a significant (pp = 0.82) sub cluster. Three sequences from Tuscany (one from Siena: SI-12520 and two from Florence: FIV-7490 and FIV-7502) significantly grouped (pp = 0.95) with a sequence from central Italy (ANV-1610-229). A second strain was found by BLAST, isolated in Sicily from an African man with unknown risk factors.

The epidemiological characteristics of subjects carrying CRF60_BC residing outside Apulia (**Table 1**) did not differ substantially from those residing in Apulia. All but one patients were males (22/23, 96%), with a median age of 25 years (range: 21–52), diagnosed from 2009 to 2014 and, for subjects with known risk factors, predominantly MSM (13/14, 92.8%). The last negative HIV-1 test, available only for three patients, suggested a recent seroconversion (less than 12 months) in two cases. However, the analysis of the fraction of ambiguous nucleotides indicated that the majority of patients (59.3%, n = 16) had been infected for more than 1 year.

Emergence of Second Generation Recombinant Forms

Three patients harbored URFs in which CRF60_BC was one of the parental strains, as indicated by the phylogenetic trees

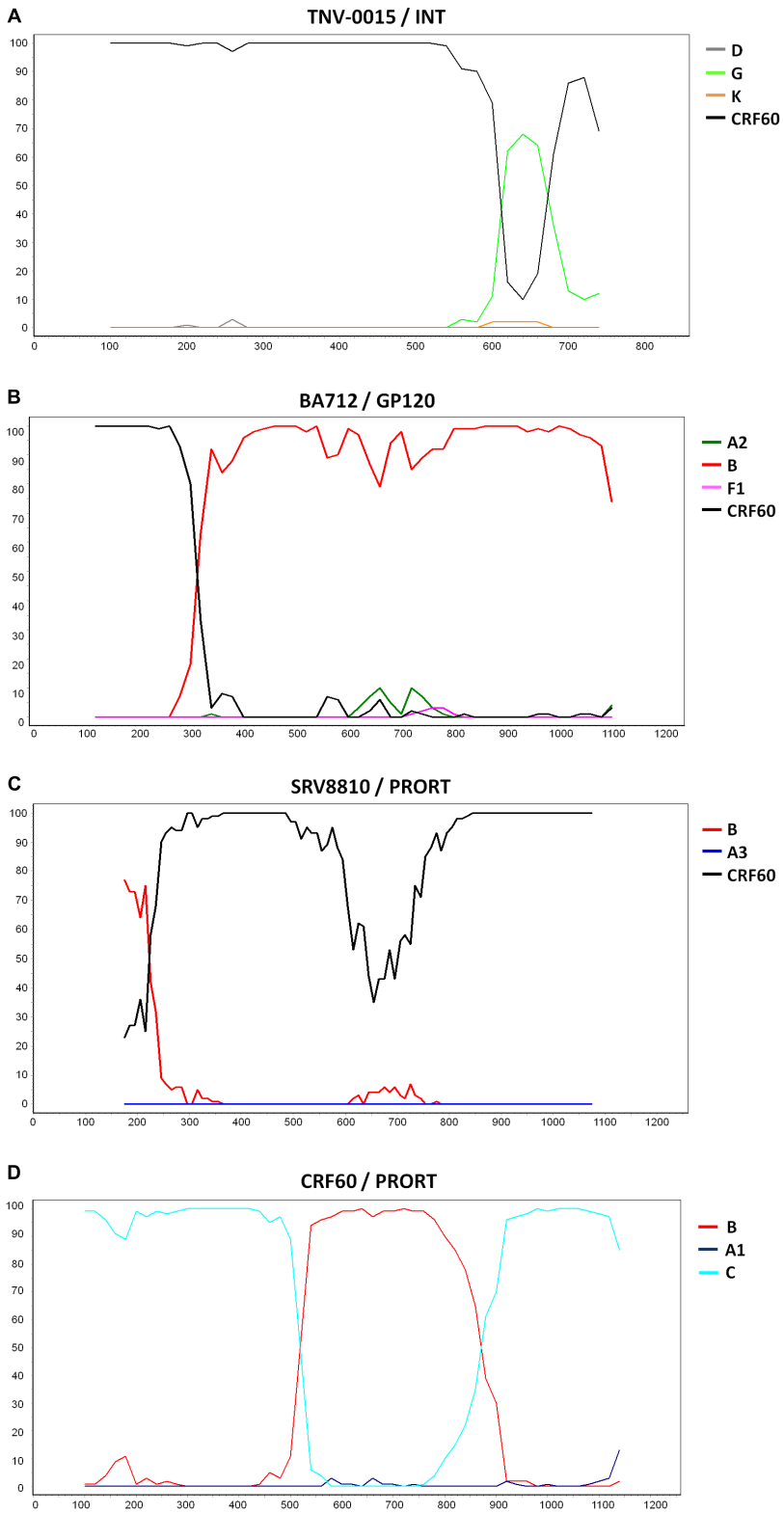


FIGURE 3 | Recombination sites of 2nd generation recombinants involving CRF60_BC. Bootscan analyses of the three unique recombinant forms (URF) in which CRF60_BC was one of the parental strains. Bootscanning plot is performed with the Simplot program using a window between 200 and 350 bp and step size between 10 and 20 bp. **(A)** Bootscan of TNV-0015 in the integrase portion. **(B)** Bootscan of BA712 in the GP120 portion. **(C)** Bootscan of SRV8810 in the PRO-RT portion. **(D)** Bootscan of CRF60_BC in the PRO-RT portion.

shown in **Supplementary Figures S1**, **S2**, **S3A**. **Figure 3** shows the bootscan analysis of these sequences and the CRF60_BC bootscan of *pol* region. Protease and RT regions of two of these sequences clustered significantly with CRF60_BC confirming the recombinant BC pattern, but different recombination patterns were detected in INT for strain TNV-0015 (**Figure 3A**) and in gp120 for BA712, in which CRF60_BC is subtype C. The bootscan analysis identified a CRF60/G mosaic pattern with two breakpoints at positions 4,848 and 4,907 (as mapped to HXB2 reference) with a 60 base-pair fragment of subtype G in the middle of the integrase region. Recombination analysis on BA712 identified a CRF60/B recombinant with one breakpoint at position 6,710 (**Figure 3B**). This recombination structure in gp120 resulted in a switch in co-receptor viral tropism from CCR5 to CXCR4. The sequence SRV8810, showed a B/CFR60 mosaicism in the PRO-RT region (**Figure 3C**) with a portion belonging to B subtype at the 3′ end of protease (position 2,489).

Supplementary Figures S1, **S2**, **S3B,C** show the maximum likelihood trees for individual segments identified within each breakpoint confirming the recombination pattern for all recombinants. Despite the presence of TNV-0015 (**Supplementary Figure S1C**) and SRV8810 (**Supplementary Figure S3B**) in subtype G and B clades respectively, these clusters are not bootstrap supported due to the limited length of the analyzed fragments.

Phylodynamic Analysis

Evolutionary population dynamics were estimated based on the data sets of non-recombinant regions (gp120 and gp41) of CRF60_BC sequences from Apulia with known data. The INT portion was not evaluated due to the limited availability of sequences. After testing different coalescent priors (constant population size, exponential growth, logistic growth, and Bayesian skyline plot), using both strict and relaxed molecular clock models, the Bayes factor (BF) analysis showed that the relaxed lognormal molecular clock fitted the data better than the strict clock model (2lnBF = 114.5) and the Bayesian skyline plot fitted the data better than the other models. The estimated mean rate of evolution was 7×10^{-3} substitutions per site per year (s/s/y) (95%HPD: 5.1×10^{-3}–9.2×10^{-3}) and 6.2×10^{-3} s/s/y (95%HPD: 3.4×10^{-3}–9.3×10^{-3}) for the gp120 and gp41, respectively. The time of the most recent common ancestor (tMRCA) corresponding to the root of the trees was estimated to be 11.4 years before the latest sampling data for both regions (95%HPD: 11–12.85 years), corresponding to mid 2004, very close to the year of diagnosis of the oldest sequence in our data set (BAV275, 2005).

The Bayesian Skyline plot (**Supplementary Figure S4**) shows the changes in population size at different times from the root of the tree to the time of the most recent samples (year 2016 and 2017). The estimated effective number of infections exponentially grew from the middle of 2006 to 2011, when the initial outbreak was identified; after this period, the curve reached a plateau and remained relatively constant until the latest calendar years.

DISCUSSION

Although the first CRFs have been found in Africa and mainly Central Africa, inter-subtype recombination events have also occurred in other continents, leading to the spread of new CRFs. Several CRFs have been described in Europe, as different non-B subtypes circulate together with the founder B clade and are detected at increasing prevalence almost in all countries (Tebit and Arts, 2011). Several European CRFs show a mosaic structure due to multiple recombination events with different parental subtypes; however, the B clade is the most frequent parental subtype. Indeed, HIV-1 superinfection, which underlies the recombination events, has been frequently reported in MSM due to the numerous risk contacts (Redd et al., 2013). As the B variant had a founder effect within intravenous drug users in some European countries including Italy (Lukashov et al., 1996), some of these CRFs (CRF03_AB and CRF14_BG) mostly segregate in this specific high risk population (Hemelaar, 2013). Some recombinant forms (i.e., CRF05_DF and CRF47_BF) have generated limited epidemics both in frequency and regional dispersion (Fernandez-Garcia et al., 2010), while others (CRF03_AB and CRF14_BG) have had a relevant impact on the HIV-1 epidemic, both in the countries of first detection and neighboring countries (Liitsola et al., 1998; Delgado et al., 2002).

We previously described a new CRF epidemic in a single Italian region (Apulia) mainly involving young MSM seeking sexual partners through on-line dating websites, who acquired HIV-1 infection in 2009–2011 and were diagnosed soon after seroconversion (Monno et al., 2012; Simonetti et al., 2014). The subtype C regions included in CRF60_BC are highly similar to the South American variant, suggesting that CRF60_BC originated in South America where subtypes B and C largely co-circulate and other BC inter-subtype recombinants have been documented (Hemelaar, 2013). Alternatively, CRF60_BC might have originated in Italy where, subtype C, mostly harbored by recent immigrants and Italian individuals who traveled to endemic regions, accounts for 13% of non-B variants (Lai et al., 2010). Additionally, our group recently conducted a phylogeographic analysis revealing that the occurrence of subtype C in Italy is related to the South American variant, likely due to transmission chains among MSM (Lai et al., 2014a). To follow-up the epidemic potential of the originally described CRF60_BC, we traced its spread through a large Italian database collecting HIV-1 sequences prospectively.

CRF60_BC was originally limited to the Apulia region but then reached distant regions of Italy. Increased sequence diversity after the initial outbreak suggests evolutionary bottlenecks along the transmission chains and adaptation to the host immune pressure. This observation is supported by the fact that some mutations, predominantly non-synonymous mutations, were fixed overtime and that the growing diversity among strains significantly correlates with the time of HIV-1 diagnosis. In this study, most CRF60_BC variants circulate among young MSM who recently acquired HIV-1 infection during the 2005–2017 period. These characteristics did not differ from those reported for patients described in the initial outbreak, indicating that CRF60_BC circulation is actually closely linked to MSM. Due to

the higher prevalence in this population and HIV-1 increasing genetic diversity, the MSM population might become a new recombination hotspot, as observed in previous studies in China reporting several CRFs among MSM population (Han et al., 2013; Wu et al., 2013; Zhang et al., 2014). Noteworthy, the number of sequences identified so far is three fold higher compared to our first report.

In addition, we detected three second-generation recombinants which further support a widespread circulation of CRF60_BC, all of which showed recombination in different portions of the HIV-1 genome. In one case, the recombination pattern involved subtype G, an uncommon subtype that is present only at 6% of non-B subtype strains (Lai et al., 2010). In another case, recombination in gp120 entailed the coreceptor switch to CXCR4, despite the recent acquisition of HIV-1 infection in 2012 confirmed by CD4+ T-cell count and viral load measure at diagnosis. By identifying new recombinant forms with a BC pattern in MSM, we demonstrated that these individuals are likely to represent a preferential group that may give rise to HIV-1 recombinants.

The finding of a patient carrying CRF60_BC diagnosed in 2005 indicates that the circulation of this strain probably began in the early 2000s. The phylodinamic analysis set the origin of CRF60_BC in the mid 2004 (range: 2003–2005) suggesting that this subject could be one of the first cases of CRF60_BC. The exponential growth of infections sustained by CRF60_BC was observed in 2006–2011 period when most patients were identified, suggesting a rapid spread among young MSM with recent infection through Italy.

This study has some limitations. First, information related to studied patients was not always available or complete. Second, the data set in the current study may not be representative of the entire country and we may have missed a number of recently infected individuals present in additional local networks. Regarding the unique recombinants, additional studies involving single genome amplification and full-length genome sequencing, are required to extend the characterization of the recombination patterns and to exclude the role of superinfection. Despite our analysis might represent an underestimate, our data strongly suggest that CRF60_BC gained epidemic importance, spreading in multiple Italian regions and increasing its population size in very few years (2007–2013), among young MSM.

This is valuable information for public health agencies developing strategies to prevent the HIV-1 epidemic to spread into a more complex epidemiological landscape. In addition, our results further highlight the need for better prevention campaigns in young MSM, who represent a population with a poor control of HIV-1 transmission (Volz et al., 2018). Finally, future studies should continue to focus on transmission clusters and pay special attention to recombinant forms, as these variants not only reflect the changing landscape of HIV-1 diversity, but can also unveil the onset of new epidemic bursts.

ETHICS STATEMENT

ARCA is an observational HIV cohort approved by the Regional Ethical Committee of Tuscany (Comitato Etico Area Vasta Toscana Sudest). The study was conducted in accordance with the 1964 Declaration of Helsinki and the ethical standards of the Italian Ministry of Health. Patients included on ARCA database signed an informed consent and agreed to have their anonymized data stored on a central server and used for research studies.

AUTHOR CONTRIBUTIONS

AL and FS conceived and designed the study. GB, GA, and LM collected and analyzed the epidemiological and viral data of patients from Apulia. SG, GS, CM, MZ, SM, and PB provided epidemiological and viral data of patients from ARCA database. AL, FS, AB, and CB wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank all patients included in the study and all centers participating to the ARCA database.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00769/full#supplementary-material

REFERENCES

Addo, M. M., Yu, X. G., Rathod, A., Cohen, D., Eldridge, R. L., Strick, D., et al. (2003). Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* 77, 2081–2092.

Allen, T. M., Altfeld, M., Geer, S. C., Kalife, E. T., Moore, C., O'sullivan, K. M., et al. (2005). Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J. Virol.* 79, 13239–13249.

Apps, R., Qi, Y., Carlson, J. M., Chen, H., Gao, X., Thomas, R., et al. (2013). Influence of HLA-C expression level on HIV control. *Science* 340, 87–91. doi: 10.1126/science.1232685

Baryshev, P. B., Bogachev, V. V., and Gashnikova, N. M. (2014). HIV-1 genetic diversity in Russia: CRF63_02A1, a new HIV type 1 genetic variant spreading in Siberia. *AIDS Res. Hum. Retroviruses* 30, 592–597. doi: 10.1089/AID.2013.0196

Delgado, E., Thomson, M. M., Villahermosa, M. L., Sierra, M., Ocampo, A., Miralles, C., et al. (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* 29, 536–543.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Fernandez-Garcia, A., Delgado, E., Cuevas, M. T., Vega, Y., Montero, V., Sanchez, M., et al. (2016). Identification of an HIV-1 BG intersubtype recombinant form (CRF73_BG), partially related to CRF14_BG, which is circulating in portugal and spain. *PLoS One* 11:e0148549. doi: 10.1371/journal.pone.0148549

Fernandez-Garcia, A., Perez-Alvarez, L., Cuevas, M. T., Delgado, E., Munoz-Nieto, M., Cilla, G., et al. (2010). Identification of a new HIV type 1 circulating BF intersubtype recombinant form (CRF47_BF) in Spain. *AIDS Res. Hum. Retroviruses* 26, 827–832. doi: 10.1089/aid.2009.0311

Foster, G. M., Ambrose, J. C., Hue, S., Delpech, V. C., Fearnhill, E., Abecasis, A. B., et al. (2014). Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of circulating recombinant form (CRF) 50_A1D. *PLoS One* 9:e83337. doi: 10.1371/journal.pone.0083337

Gonzalez-Domenech, C. M., Viciana, I., Delaye, L., Mayorga, M. L., Palacios, R., De La Torre, J., et al. (2018). Emergence as an outbreak of the HIV-1 CRF19_cpx variant in treatment-naive patients in southern Spain. *PLoS One* 13:e0190544. doi: 10.1371/journal.pone.0190544

Han, X., An, M., Zhang, W., Cai, W., Chen, X., Takebe, Y., et al. (2013). Genome sequences of a Novel HIV-1 circulating recombinant form, CRF55_01B, identified in China. *Genome Announc.* 1:e50–12. doi: 10.1128/genomeA.00050-12

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Hemelaar, J. (2013). Implications of HIV diversity for the HIV-1 pandemic. *J. Infect.* 66, 391–400. doi: 10.1016/j.jinf.2012.10.026

Hu, W. S., and Hughes, S. H. (2012). HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* 2:a006882. doi: 10.1101/cshperspect.a006882

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., et al. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7552–7557. doi: 10.1073/pnas.0802203105

Kouri, V., Aleman, Y., Perez, L., Perez, J., Fonseca, C., Correa, C., et al. (2014). High frequency of antiviral drug resistance and non-b subtypes in HIV-1 patients failing antiviral therapy in Cuba. *J. Int. AIDS Soc.* 17:19754. doi: 10.7448/IAS.17.4.19754

Kouri, V., Khouri, R., Aleman, Y., Abrahantes, Y., Vercauteren, J., Pineda-Pena, A. C., et al. (2015). CRF19_cpx is an evolutionary fit HIV-1 variant strongly associated with rapid progression to AIDS in Cuba. *EBioMedicine* 2, 244–254. doi: 10.1016/j.ebiom.2015.01.015

Kouyos, R. D., Von Wyl, V., Yerly, S., Boni, J., Rieder, P., Joos, B., et al. (2011). Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. *Clin. Infect. Dis.* 52, 532–539. doi: 10.1093/cid/ciq164

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lai, A., Bozzi, G., Franzetti, M., Binda, F., Simonetti, F. R., De Luca, A., et al. (2016). 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. *PLoS One* 11:e0146097. doi: 10.1371/journal.pone.0146097

Lai, A., Bozzi, G., Franzetti, M., Binda, F., Simonetti, F. R., Micheli, V., et al. (2014a). 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. doi: 10.1002/jmv.23891

Lai, A., Ciccozzi, M., Franzetti, M., Simonetti, F. R., Bozzi, G., Binda, F., et al. (2014b). Local and global spatio-temporal dynamics of HIV-1 subtype F1. *J. Med. Virol.* 86, 186–192. doi: 10.1002/jmv.23783

Lai, A., Riva, C., Marconi, A., Balestrieri, M., Razzolini, F., Meini, G., et al. (2010). Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. *HIV Med.* 11, 593–602. doi: 10.1111/j.1468-1293.2010.00832.x

Lai, A., Simonetti, F. R., Zehender, G., De Luca, A., Micheli, V., Meraviglia, P., 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. doi: 10.1371/journal.pone.0042223

Leoz, M., Feyertag, F., Charpentier, C., Delaugerre, C., Wirden, M., Lemee, V., et al. (2013). Characterization of CRF56_cpx, a new circulating B/CRF02/G recombinant form identified in MSM in France. *AIDS* 27, 2309–2312. doi: 10.1097/QAD.0b013e3283632e0c

Liitsola, K., Tashkinova, I., Laukkanen, T., Korovina, G., Smolskaja, T., Momot, O., 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.

Lukashov, V. V., Huismans, R., Rakhmanova, A. G., Lisitsina, Z. N., Akhtyrskaya, N. A., Vlasov, N. N., 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.

Lukashov, V. V., Kuiken, C. L., Vlahov, D., Coutinho, R. A., and Goudsmit, J. (1996). Evidence for HIV type 1 strains of U.S. Intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. *AIDS Res. Hum. Retroviruses* 12, 1179–1183.

Monno, L., Brindicci, G., Lai, A., Punzi, G., Altamura, M., Simonetti, F. R., et al. (2012). An outbreak of HIV-1 BC recombinants in Southern Italy. *J. Clin. Virol.* 55, 370–373. doi: 10.1016/j.jcv.2012.08.014

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B., and 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.

Peeters, M., D'arc, M., and Delaporte, E. (2014). Origin and diversity of human retroviruses. *AIDS Rev.* 16, 23–34.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Rawson, J. M. O., Nikolaitchik, O. A., Keele, B. F., Pathak, V. K., and Hu, W. S. (2018). Recombination is required for efficient HIV-1 replication and the maintenance of viral genome integrity. *Nucleic Acids Res.* 46, 10535–10545. doi: 10.1093/nar/gky910

Recordon-Pinson, P., Alves, B. M., Tumiotto, C., Bellecave, P., Bonnet, F., Neau, D., et al. (2018). A new HIV-1 circulating recombinant form (CRF98_cpx) between CRF06_cpx and subtype B Identified in southwestern france. *AIDS Res. Hum. Retroviruses* 34, 1005–1009. doi: 10.1089/AID.2018.0122

Redd, A. D., Quinn, T. C., and Tobian, A. A. (2013). Frequency and implications of HIV superinfection. *Lancet Infect. Dis.* 13, 622–628. doi: 10.1016/S1473-3099(13)70066-5

Robertson, D. L., Sharp, P. M., Mccutchan, F. E., and Hahn, B. H. (1995). Recombination in HIV-1. *Nature* 374, 124–126.

Rossetti, B., Di Giambenedetto, S., Torti, C., Postorino, M. C., Punzi, G., Saladini, F., et al. (2018). Evolution of transmitted HIV-1 drug resistance and viral subtypes circulation in Italy from 2006 to 2016. *HIV Med.* 19, 619–628. doi: 10.1111/hiv.12640

Simonetti, F. R., Lai, A., Monno, L., Binda, F., Brindicci, G., Punzi, G., et al. (2014). Identification of a new HIV-1 BC circulating recombinant form (CRF60_BC) in Italian young men having sex with men. *Infect. Genet. Evol.* 23, 176–181. doi: 10.1016/j.meegid.2014.02.007

Struck, D., Roman, F., De Landtsheer, S., Servais, J. Y., Lambert, C., Masquelier, C., et al. (2015). Near full-length characterization and population dynamics of the human immunodeficiency virus type I circulating recombinant form 42 (CRF42_BF) in luxembourg. *AIDS Res. Hum. Retroviruses* 31, 554–558. doi: 10.1089/AID.2014.0364

Tebit, D. M., and Arts, E. J. (2011). Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect. Dis.* 11, 45–56. doi: 10.1016/S1473-3099(10)70186-9

Thomson, M. M., Perez-Alvarez, L., and Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* 2, 461–471.

Volz, E. M., Le Vu, S., Ratmann, O., Tostevin, A., Dunn, D., Orkin, C., et al. (2018). Molecular epidemiology of HIV-1 subtype B reveals heterogeneous transmission risk: implications for intervention and control. *J. Infect. Dis.* 217, 1522–1529. doi: 10.1093/infdis/jiy044

Wu, J., Meng, Z., Xu, J., Lei, Y., Jin, L., Zhong, P., et al. (2013). New emerging recombinant HIV-1 strains and close transmission linkage of HIV-1 strains in the Chinese MSM population indicate a new epidemic risk. *PLoS One* 8:e54322. doi: 10.1371/journal.pone.0054322

Zhang, W., Han, X., An, M., Zhao, B., Hu, Q., Chu, Z., et al. (2014). Identification and characterization of a novel HIV-1 circulating recombinant form (CRF59_01B) identified among men-who-have-sex-with-men in China. *PLoS One* 9:e99693. doi: 10.1371/journal.pone.0099693

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MC declared past co-authorships with two of the authors AL and MG to the handling Editor.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 May 2019
doi: 10.3389/fmicb.2019.01006

Characterization of New Recombinant Forms of HIV-1 From the Comunitat Valenciana (Spain) by Phylogenetic Incongruence

Beatriz Beamud[1,2], *María Alma Bracho*[2,3,4] *and Fernando González-Candelas*[1,2,4]*

[1] Instituto de Biología Integrativa de Sistemas, Consejo Superior de Investigaciones Científicas, Universitat de València, Valencia, Spain, [2] Unidad Mixta de Investigación Infección y Salud Pública, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana, Universitat de València, Valencia, Spain, [3] Área de Genómica y Salud, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana, Valencia, Spain, [4] Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Valencia, Spain

Recombination is one of the main processes shaping the evolution of HIV-1, with relevant consequences for its epidemiology. In fact, Circulating and Unique Recombinant Forms (CRFs and URFs) cause 23% of current infections. The routine analyses of antiretroviral resistance yield partial *pol* gene sequences that can be exploited for molecular epidemiology surveillance but also to study viral diversity and to detect potential recombinant samples. Among the *pol* sequences derived from a large sample dataset from the Comunitat Valenciana (Spain), we identified nine putative recombinant samples. We aimed at fully characterizing these samples and performing a detailed analysis of the corresponding recombination events. We obtained nearly full-genome sequences and used jpHMM and RDP4 to detect and characterize recombinant fragments. We assessed the confidence of these inferences by likelihood mapping and phylogenetic placement with topology congruence tests. Next, we performed a phylogenetic analysis of each putative recombinant fragment to determine its relationships to previously described recombinant forms. We found that two samples related to CRF44_BF whereas the rest corresponded to new URFs (two URF_AD, one URF_BG that can constitute a new CRF resulting from subtype B and CRF24_BG, and two URF_cpx composed of A, G, K, H, and J subtypes). These URFs have a complex recombination pattern that cannot be determined accurately. They seem to have arisen by successive recombination events among lineages, including other CRFs. Our results highlight the usefulness of routine surveillance analysis for the detection of new HIV-1 recombination forms and, at the same time, the need for full-genome sequencing and recombination detection guidelines to properly characterize this complex process.

Keywords: HIV-1, nearly-full genome, recombination, phylogeny, CRFs, URFs

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Miguel Arenas,
University of Vigo, Spain
Federico Garcia,
Hospital Universitario San Cecilio, Spain

***Correspondence:**
Fernando González-Candelas
fernando.gonzalez@uv.es

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 October 2018*
Accepted: *18 April 2019*
Published: *22 May 2019*

Citation:
Beamud B, Bracho MA and González-Candelas F (2019) Characterization of New Recombinant Forms of HIV-1 From the Comunitat Valenciana (Spain) by Phylogenetic Incongruence. Front. Microbiol. 10:1006. doi: 10.3389/fmicb.2019.01006

INTRODUCTION

The high genetic variability of HIV-1 is one of the main hurdles to control the current epidemic of this virus. This diversity is the result of high mutation (Abram et al., 2010), replication (Rodrigo et al., 1999), and recombination rates. The recombinogenic nature of HIV-1 is well-known, as this virus has one of the highest rates known (Rodrigo et al., 1999; Shriner et al., 2004). Recombination

has important evolutionary consequences for viruses and has been associated with the expansion of viral host ranges, the emergence of new variants, increases in pathogenesis and virulence, the alteration of tropisms, the immune escape, and resistance to antivirals (Martin et al., 2011; Simon-Loriere and Holmes, 2011). In addition, recombination also plays an important role in HIV-1 epidemiology. Up to date, 91 circulating recombinant forms (CRFs) have been described (Los Alamos National Laboratory[1]). CRFs are recombinant HIV-1 genomes that share the same recombination breakpoints (BPs) between the same parentals and have been found in at least three non-epidemiologically related individuals. If a recombinant form does not fulfill the requirements to be considered a CRF, it is called a unique recombinant form (URF). RFs can also be classified according to their mosaic structure: they are denoted as complex (cpx) if they are composed by three or more subtypes or second-generation recombinants (SGRs) if they result from a recombination event in which pre-existing RFs are involved. RFs account for 23% of the current infections and are responsible for almost half of total infections in areas such as Central Africa (Hemelaar et al., 2019).

HIV-1 subtyping is still dominated by the partial sequencing of the *pol* gene used in routine analyses of antiretroviral resistance mutations. However, the development of whole genome sequencing (WGS) and its reducing costs have led to a substantial increase in the availability of complete HIV-1 genomes. This is partially responsible for the increased reporting of RFs world-wide. This increase also results from the mixing of newly formed HIV-1 variants, resulting in a complex and dynamic epidemiology (Hemelaar et al., 2019). Spain is one of the western European countries with the highest prevalence of HIV-1[2]. Subtype B is responsible for 82–88% of the infections in this country followed by RFs, which cause about 9–10% of cases (Yebra et al., 2012). In fact, the three CRFs described from Spanish samples so far, CRF14_BG, CRF47_BF, and CRF73_BG (Delgado et al., 2002; Fernández-García et al., 2010, 2016), involve the B subtype. Apart from these, only a few studies with complete HIV-1 genome sequences have been done in this country using non-B subtypes (Casado et al., 2003; Sierra et al., 2005; Holguin et al., 2008; Cuevas et al., 2010). Hence, the prevalence of recombinant forms in Spain, as in many other countries, might be underestimated (Yebra et al., 2012).

The *pol* sequences obtained in the analyses of resistance mutations can be used to gain insight into the population dynamics and evolution of the corresponding lineages. These analyses can also identify sequences that are not neatly included in any of the major subtypes or sub-subtypes of HIV-1. They may represent new variants or RFs, relevant for the epidemiological characterization of the viral population. There is no standard procedure to detect recombination in HIV-1 and different methods have been used so far (a comprehensive list of recombination analysis software[3]). Due to this lack of standardization and the intrinsic difficulties in identifying recombinant events, some authors have postulated that several methods should be used to obtain a precise picture of recombination (Posada, 2002). The hallmark of recombination is the presence of fragments with different ancestries in the same genome. Therefore, the ultimate test for recombination should be based on revealing a statistically significant lack of phylogenetic congruence among portions of the recombinant genomes (Pérez-Losada et al., 2015). In order to test the statistical significance of alternate topologies, the corresponding multiple alignments must incorporate adequate levels of genetic variation. Consequently, analyses of the phylogenetic signal should also be included in this procedure. This is most relevant for small recombinant regions in which an insufficient phylogenetic signal may lead to wrong inferences of phylogenies (Jia et al., 2014).

During routine analyses of resistance mutations in HIV-1 samples from the Comunitat Valenciana (Spain) (Patiño-Galindo et al., 2017), we detected several samples whose partial *pol* sequences did not cluster with high support in any sub-subtype. Our aim here is to describe their full genomic characterization and to perform an in-depth analysis of recombination using different methods and confirm them by means of phylogenetic congruence tests.

MATERIALS AND METHODS

Samples

From 2010 to 2013, about 1,800 serum samples of individuals infected with HIV-1 were screened for antiretroviral resistance mutations in our laboratory (Patiño-Galindo et al., 2017). For this purpose, a partial region (1,302 nt) of the *pol* gene was sequenced. The phylogenetic analysis of this region revealed nine samples with a suspected recombinant nature because they did not cluster with high enough bootstrap support with reference sequences of pure subtypes and representative CRFs (data not shown). These nine samples, all from Spanish patients that attended hospitals of the Comunitat Valenciana (Spain), were subjected to nearly-WGS.

Extraction, Amplification, and Sequencing

RNA was extracted from sera using the NucliSENS® EasyMag® automated platform (bioMérieux, Marcy, L'Étoile, France). Reverse transcription (RT) of RNA was performed in two adjacent genome regions to improve yield. The first hemigenome was obtained by priming with IN-T-A2 (Thomson et al., 2002) in the *vif* gene (nts: 5192–5218, HXB2, GenBank Accession No. K03455). The second hemigenome was obtained with UNINEF7′ (Nadai et al., 2008), close to the 3′ end of the viral RNA (nts: 9605–9632, HXB2). The conditions for RT and PCR were slightly modified from Nadai et al. (2008). RNA (4 μL) was reverse transcribed in a total volume of 20 μL with 500 uM dNTP, 2.5 uM primer, 1x RT buffer, 40 U RNasin® (Promega, Madison, WI, United States), 10 mM DTT and 200 U of SuperScript™ III reverse transcriptase (Invitrogen™, Carlsbad, CA, United States). Reagents and RNA were incubated for cDNA synthesis at 50°C for 2 h, followed by 85°C for 5 min.

[1]http://www.hiv.lanl.gov

[2]http://apps.who.int/gho/data/node.main.620?lang=en

[3]http://bioinf.man.ac.uk/robertson/recombination/programs.shtml

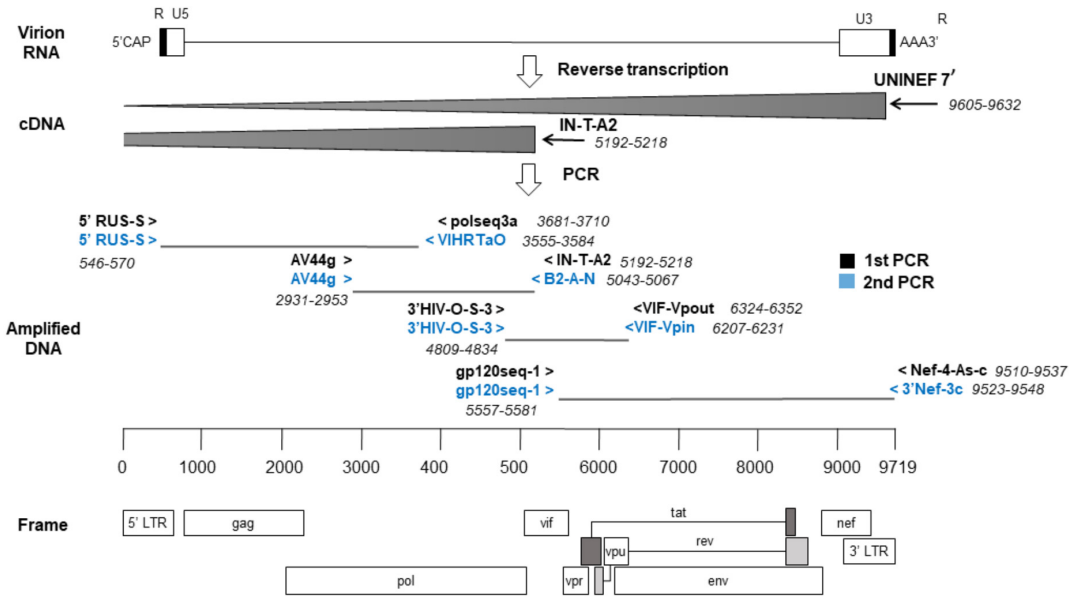


FIGURE 1 | Sequencing strategy by the nearly full-length RT-PCR method. The locations of the primers used are relative to the reference genome, HXB2. cDNA synthesis was done in two reactions and a strategy of four overlapping amplicons was followed by PCR.

Next, four overlapping regions were amplified using first-round PCRs and, if needed, nested-PCRs (**Figure 1**). The first amplicon consisted of about 3 kb from 5′ LTR U5 to the middle of *pol* (nts: 571–3554, HXB2). The second amplicon, of 2.1 kb, covered from the second half of *pol* to the start of *vif* (nts: 2930–5042, HXB2). The third amplicon, of approximately 1.4 kb, covered until *env* gp160. Lastly, the fourth amplicon consisted of 3.9 kb from *env* gp160 to the end of *nef* (nts: 5582–9511, HXB2). The reaction volume was 25 μL, containing 1x PCR buffer, 350 uM dNTP mixture, 0.4 uM of each primer and 2.5 U TaKaRa Ex Taq® (Takara Bio, Shiga, Japan). Cycling conditions were 94°C for 2 min and then 10 cycles of (94°C 10 s, 60°C 30 s, 68°C 3 min), and 20 cycles of (94°C 10 s, 55°C 30 s, 68°C 3 min) followed by 68°C for 10 min. Nested-PCRs were done with the same conditions as above but adding 1.5 μL of the previous PCR product with appropriate primers.

Positive PCR products were purified with a High Pure PCR Product Purification Kit (Roche Applied Science, Indianapolis, IN, United States) and sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, United States) in an ABI 3730xl DNA sequencer (Roche, Nutley, NJ, United States). The set of primers used in this study is available upon request. We obtained the consensus nearly full-length genome (8.9 kb) by assembling the sequence readings that passed the default quality filter of Pregap4 (Bonfield et al., 1995; Nadai et al., 2008) along with visual inspection of polymorphic regions. For genomes with more than one contig, scaffolding was done after comparison with the closest reference genome available at the Los Alamos National Laboratory (LANL) database. The final sequences have been deposited in the GenBank with Accession Nos. MK095228–MK095236.

Recombination Analysis

The mosaic recombinant structure of the samples was screened using two different strategies. Firstly, breakpoints (BPs) were determined with the default jpHMM options (Schultz et al., 2010). Secondly, we obtained recombination events and BPs that were supported by at least three of the seven selected tests implemented in RDP4: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), BootScan (Martin et al., 2005), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001), SiScan (Gibbs et al., 2000) and 3Seq (Boni et al., 2007). Default parameter values were used except for window size that was increased to 60 nt in RDP, to 120 in MaxChi and Chimaera, and to 500 in BootScan and SiScan, as suggested (Lemey et al., 2009).

For jpHMM we used the nine sequences obtained above. To avoid confounding effects between recombinant samples, we used a separate alignment for each sample in RDP4. These alignments consisted of one sequence of each 'pure' subtype and sub-subtypes A1, A2, F1, and F2 (from RIP Alignment[4]) and each putative recombinant sample using the option -add of MAFFT v.7 (Katoh and Standley, 2013).

After detecting recombination, we wanted to further contrast the confidence of each recombination event independently. For this, we used a slightly modified phylogenetic congruence testing pipeline. The corresponding BPs of each detection strategy delimited sequence regions that were extracted and independently aligned using MAFFT v.7 (Katoh and Standley, 2013) *(option –add)* with the RIP reference alignment detailed above. This partial alignment was used to reconstruct a reference maximum likelihood (ML) tree using IQ-TREE

[4]https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html#RIP

TABLE 1 | Summary statistics of the near complete genomes obtained for the 9 HIV-1 samples studied.

Sample	Total reads	Mean length of reads (nt)	Contigs	Ambiguities (N's)	Assembly length (nt)	HXB2 coverage	Depth of coverage
79	55	740.49	1	25	8987	0.92	4.53
93	60	540.03	2	354	8922	0.87	3.66
678	10	665.30	1	4	4016	0.41	1.66
703	47	586.17	1	0	8994	0.91	3.06
724	53	558.04	1	84	9012	0.90	3.29
2011	54	613.93	2	396	8852	0.86	3.92
2104	35	599.49	2	352	8934	0.88	2.44
3011	33	694.88	1	0	8961	0.91	2.56
3164	9	883.00	2	2538	7379	0.49	1.64

v.1.6.5 (Nguyen et al., 2015) with the best evolutionary model (GTR+F+I+G4) determined by the same software (Kalyaanamoorthy et al., 2017). To test the phylogenetic congruence of the recombinant fragments in each genome, we constructed all possible alternative trees by clustering each sample with all the (sub-)subtypes in the reference alignment. For this, we used the R packages ape (Paradis and Schliep, 2018) and phytools (Revell, 2011). The 11 alternative trees for each sample were used for Expected Likelihood Weights (ELWs) (Strimmer and Rambaut, 2002) topological congruence test with the partial alignment corresponding to each fragment. This test assigns relative support values to the alternative topologies based on the likelihood values of the corresponding alignments. These tests were performed with IQ-TREE v.1.6.5 using 10,000 RELL replicates. Additionally, the level of phylogenetic signal of each partial alignment was also assessed by means of likelihood mapping with the evaluation of 1,000 random quartets using IQ-TREE v.1.6.5. A proportion higher than 0.25 of unresolved quartets was considered to represent a lack of phylogenetic signal (Strimmer and Rambaut, 2002; Lemey et al., 2009).

Next, each fragment of the recombinant samples was further characterized by phylogenetic analysis with representative sequences of HIV-1, including RFs, as above. The alignment was obtained from LANL database (RIP v.2017, 145 sequences) but adding, with MAFFT v.7, all the URFs identified as such (resulting in a total of 226 sequences). Then, new ML trees were built with 1,000 bootstrap replicates. The closest relatives (CRs) of each fragment were determined considering as CRs the minimum set of sequences that grouped with the fragment of interest with a bootstrap support value higher than 70%.

Finally, the inferred mosaic structure from the above analyses was drawn using the ggplot2 library in R (Wickham, 2016). These analyses, except the execution of RDP4, have been integrated into a publicly available pipeline accessible at https://github.com/BBeamud/PhyloRecomb/.

RESULTS

We obtained nearly complete genome sequences of seven samples, but our sequencing strategy failed to yield enough information in two cases. Between 86 and 92% of the reference genome (HXB2) length was covered in our samples. The per base coverage ranged from 2.44 to 4.53. In four of these genomes, it was not possible to assemble all the sequencing reads into a single contig due to hypervariability in the *env* gene. This is reflected in the proportion of undetermined bases of each sequence that were used for scaffolding (**Table 1**). For the incomplete genomes, we obtained 41% (sample 678) and 49% (sample 3164) of the reference genome length with coverage depths of 1.64 and 1.66, respectively. Additional sequencing of these two samples was not possible due to lack of biological material.

Recombinant mosaic structures were obtained by two strategies. Firstly, we considered the recombination events and their defining BPs supported by at least three of the methods tested in RDP4. Additionally, we considered the BPs reported by jpHMM. These two programs assign a subtype for each fragment. All the strategies revealed putative recombination events in all the samples, but the total number of such events differed among them. To facilitate the comparison between detection strategies, we retained as major parental (MP) the subtype with most genomic nucleotides assigned. In addition, we considered the same recombination event if the minor assigned subtype and the inferred BPs were similar ($\pm$ 200 nt) in both sides for both strategies. A total of 42 recombination events were detected but only eight were shared by the two strategies. Of the 34 remaining events, 21 were detected exclusively by jpHMM and 13 by RDP4. There were substantial differences in the performance of the methods implemented in RDP4, with a minimum of 3 (GENECONV) and a maximum of 19 (MaxChi and 3Seq) detected events (**Table 2**). Shared events had an average length of 1454.9 nt with a difference in the start and end positions of 14.9 and 59.0 nt, respectively. The shared events detected by RDP4 were on average 11.9 nt shorter than those detected by jpHMM. On the contrary, unique events by jpHMM and RDP4 had lengths of 513.6 and 1399.4 nt. Surprisingly, some large events detected by RDP4 were not shared with jpHMM. This was due to RDP4 failing to detect several events of the minor subtypes and reporting them as unique, larger recombination events (**Table 3**, samples 2011, 2104, 3011, and 3164).

To further corroborate these events, we performed ELW congruence tests and likelihood mapping analyses for each fragment. A recombination event was considered verified only when the tree with the sample fragment grouped with the minor subtype identified had a relative support value, as determined by ELW, larger than 0.5 than that of the second-best tree.

Additionally, we considered fragments with less than 0.25 unresolved quartets reliable for phylogenetic inferences. All the events jointly detected by jpHMM and RPD4 were verified by ELW tests and phylogenetic signal. However, 15 out of the 21 events detected only by jpHMM were corroborated, regardless of their phylogenetic signal, whereas only 3 of the 13 events detected only by RDP4 were verified. For the latter, the main discrepancies were observed in the assigned subtype whereas for jpHMM a high proportion of cases with low phylogenetic signal was observed (**Table 3**). A summary of true positive, false positive and false negative recombination events, detected by the different methods implemented in RDP4 and jpHMM, is shown in **Figure 2**. The overall best performance corresponds to jpHMM, with 23 corroborated events of the 29 detected with this method. The worst performing method was GENECONV with only two corroborated events of the three detected.

Different mosaic structures were observed in the samples considering verified events. In particular, we observed two BF recombinants, two AD recombinants, three BG recombinants, and two complex recombinant forms. These mosaic structures were refined with the reconstruction of ML trees for each fragment based on the analysis of closest relatives (CRs) with 226 representative sequences of HIV-1. We performed these analyses for all the major and minor parental fragments although the results of the former analyses are not shown for simplification (**Table 3**).

The two BF recombinants corresponded to samples 79 and 93. They showed the same pattern and similar BPs by both jpHMM and RDP4. A single recombination event of subtype B was detected and verified in the region located between nucleotides 2,466 and 3,708 (all the positions reported are referred to HXB2). This structure resembles that of CRF38_BF (Ruchansky et al., 2009) although the CRs analyses of the recombinant fragments of samples 79 and 93 showed that they are related to CRF44_BF (Delgado et al., 2010) (**Table 3**). The latter has three recombination events (all of subtype B) that were not detected by jpHMM nor RDP4. We re-analyzed samples 79 and 93 by performing congruence analyses with the reference breakpoints of CRF44_BF[5] to confirm it relatedness. The results of the ELW tests showed a lack of congruence for the three recombination events with their grouping with subtype B, thus corroborating them (**Figure 3**). Hence, the CRs analyses allowed us to classify samples 79 and 93 as CRF44_BF, despite their initial identification as CRF38_BF.

Samples 678, 703, and 724 are BG recombinants, with G being the major parental with a single B recombination event. A similar situation to that in BF recombinants was found in this group, except for two recombination events detected only by RDP4, 703_2R (U|D) and 724_2R(H), which were not verified by ELW and phylogenetic signal tests. CRs analyses showed that the G fragment was related to CRF24_BG whereas the B fragment grouped consistently with pure B sequences or CRF51_01B (**Table 3**). To confirm that samples 678, 703, and 724 were indeed CRF24_BG we proceeded as above, with the BPs defined at the LANL database for this CRF. In this case, the results of the ELW tests did not reveal four different B recombination events, as

[5]https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/breakpoints.html#CRF44

TABLE 2 | Support of each putative recombination event detected by the different methods implemented in RDP4.

Sample (MP)[a]	Event	Subtype	Start	End	RDP	GENECONV	Bootscan	Maxchi	Chimera	SiSscan	3Seq
79 (F1)	1**	B	2470	3707	2.23E-09	NS	3.21E-11	8.77E-07	3.87E-07	NS	4.88E-14
93 (F1)	1**	B	2471	3708	5.38E-09	NS	1.92E-11	3.26E-07	7.95E-08	NS	4.88E-14
678 (G)	1**	B	2592	4802	3.60E-16	NS	1.31E-16	6.21E-19	5.28E-22	2.98E-12	5.36E-39
703 (G)	1**	B	2558	4803	1.13E-38	8.23E-04	8.96E-31	1.92E-17	7.25E-21	1.34E-15	4.88E-14
	2R	U(D)	8697	8860	NS	NS	NS	9.82E-06	2.91E-03	NS	2.07E-02
724 (G)	1**	B	2559	4804	1.29E-39	1.11E-03	1.54E-40	1.29E-18	6.97E-22	3.22E-17	4.88E-14
	2R	H	7337	7612	1.75E-03	1.77E-03	NS	3.70E-03	8.03E-04	NS	NS
2011(A1)	1**	D	2161	3069	6.96E-08	NS	9.57E-04	2.12E-04	1.78E-07	NS	6.31E-10
	2R	U(F)	3070	5923	4.30E-02	NS	NS	6.10E-04	9.32E-03	5.03E-10	2.44E-14
	3R**	D	6077	9384	6.13E-06	NS	2.42E-08	1.05E-05	7.31E-10	7.91E-09	8.30E-13
2104 (A1)	1**	D	2135	3252	7.84E-09	NS	3.86E-03	3.16E-08	5.42E-09	NS	7.25E-12
	2R	U(F2)	3253	5924	NS	NS	NS	1.36E-05	2.66E-02	1.41E-11	2.44E-14
	3R**	D	6078	9384	5.51E-04	NS	1.25E-05	1.02E-12	5.99E-12	9.09E-08	2.44E-14
3011 (A1)	1R	C	1526	1989	1.47E-02	NS	1.58E-02	6.05E-03	NS	NS	1.43E-02
	2R	U(F1)	2169	3238	NS	NS	NS	1.34E-02	4.64E-02	NS	1.49E-05
	3R	F2	3449	5662	NS	NS	NS	8.89E-03	7.98E-03	NS	5.90E-05
	7**	K	6109	6486	1.19E-05	NS	1.96E-05	1.72E-03	4.07E-04	NS	4.56E-02
	4R	U(C)	6608	6847	2.39E-02	NS	NS	NS	NS	4.11E-04	7.59E-03
	5R	U(F2)	8335	9385	NS	NS	NS	1.03E-04	1.22E-05	4.29E-09	NS
3164 (A1)	1R*	K	2727	3031	7.20E-04	NS	1.51E-03	NS	NS	NS	2.70E-03
	2R	F2	6079	6345	1.74E-07	NS	2.16E-05	1.32E-04	1.92E-05	NS	3.69E-05

*Asterisks indicate events confirmed by ELW (best topology) and phylogenetic signal (unresolved quartets, UQ < 0.25). *Supported by ELW; **Supported by ELW + Phylogenetic signal. [a]Major parental.*

TABLE 3 | Recombination events detected by jpHMM and RPD4 and phylogenetic analysis of the studied samples.

					jpHMM		
Sample (MP)[a]	**Event**	**Subtype**	**Start**	**End**	**ELW**	**UQ**	**# CRs (taxa)**
79 (F1)	1**	B	2466	3700	B	0.067	1 (44_BF)
93(F1)	1**	B	2467	3701	B	0.068	1 (44_BF)
678 (G)	1**	B	2576	4830	B	0.037	1 (B)
703(G)	1**	B	2575	4829	B	0.032	1(B)
724(G)	1**	B	2576	4830	B	0.027	1 (B)
2011 (A1)	1**	D	2135	3254	D|b	0.115	5 (D)
	2J**	D	6077	7317	D	0.044	4 (D)
	3J**	D	7602	8821	D	0.064	20 (D)
	4J	B	8822	9199	B|D|H|A2	0.109	19 (D)
	5J**	D	9200	9411	D|b	0.072	12 (D)
2104 (A1)	1**	D	2137	3255	D|b	0.123	1 (19_cpx)
	2J**	D	6079	8818	D	0.018	4 (D)
	3J	B	8820	9200	B|H|D|A2	0.117	1 (03_AB)
	4J**	D	9202	9412	D|b	0.078	14 (D)
3011 (A1)	1J**	G	1628	2287	G	0.139	1 (13_cpx)
	2J	J	2691	3116	J|K	0.252	2 (27|13_cpx)
	3J*	H	3118	3224	H|a1|a2	0.553	4 (H)
	4J	B	3483	4147	K|F2|F1|G	0.270	1 (04_cpx)
	5J*	G	4149	4607	G	0.337	1 (27_cpx)
	6J	B	5085	5706	K|J	0.104	3 (URF_U, 09_cpx, URF_0209)
	7**	K	6076	6296	K	0.149	1 (K)
	8J*	H	8335	8558	H	0.301	1 (04_cpx)
	9J*	J	8560	8670	J|c|k	0.431	1 (URF_A1D)
	10J*	H	8867	9018	H	0.269	H
3164 (A1)	1J**	K	6079	6298	K	0.156	1 (K)
	2J*	H	8336	8559	H	0.284	1 (04_cpx)
	3J*	J	8561	8663	J|c	0.456	2 (45_cpx, URF_U)
	4J*	H	8870	8998	H	0.376	4 (H)
	5J	B	9000	9299	H|B|F1|F2|J|K|G|D	0.149	222 (HIV-1)
					RDP4		
Sample (MP)[a]	**Event**	**Subtype**	**Start**	**End**	**ELW**	**UQ**	**# CRs (taxa)**
79 (F1)	1**	B	2470	3707	B	0.057	1 (44_BF)
93 (F1)	1**	B	2471	3708	B	0.067	1 (44_BF)
678 (G)	1**	B	2592	4802	B	0.026	1 (51_01B)
703 (G)	1**	B	2558	4803	B	0.028	1 (B)
	2R	U(D)	8697	8860	J|K|B|D|H|C	0.388	3 (24|23|20_BG)
724 (G)	1**	B	2559	4804	B	0.034	222 (HIV-1)
	2R	H	7337	7612	H|G|J	0.281	3 (24|23|20_BG)
2011(A1)	1**	D	2161	3069	D|b	0.147	5 (D)
	2R	U(F)	3070	5923	A1	0.013	1 (A3)
	3R**	D	6077	9384	D	0	11 (D)
2104 (A1)	1**	D	2135	3252	D|b	0.154	5 (D)
	2R	U(F2)	3253	5924	A1	0.02	1 (A3)
	3R**	D	6078	9384	D	0	17 (D)
3011 (A1)	1R	C	1526	1989	G|C	0.195	2 (32_06A1, 06_cpx)
	2R	U(F1)	2169	3238	H|K|J	0.137	1 (27_cpx)
	3R	F2	3449	5662	K|C|J	0.034	1 (27_cpx)
	7**	K	6109	6486	K	0.125	1 (39_BF)
	4R	U(C)	6608	6847	A2|A1|H|F2	0.396	1 (URF_U)
	5R	U(F2)	8335	9385	H	0.026	5 (H)
3164 (A1)	1R*	K	2727	3031	K	0.307	1 (49_cpx)
	2R	F2	6079	6345	K	0.172	2 (K)

*Events detected by jpHMM and RDP4 are named with one single number and those detected by only one method are named with a number and a letter (J = jpHMM, R = RDP4). Asterisks indicate events confirmed by ELW (best topology) and phylogenetic signal (unresolved quartets, UQ < 0.25). Alternative topologies found by ELW (less than 0.5 difference with the best topology) are shown in lowercase. The number of closest relatives (CRs) and corresponding pure subtype or RF are shown in the last column. [a]Major parental; *Supported by ELW; **Supported by ELW + Phylogenetic signal.*

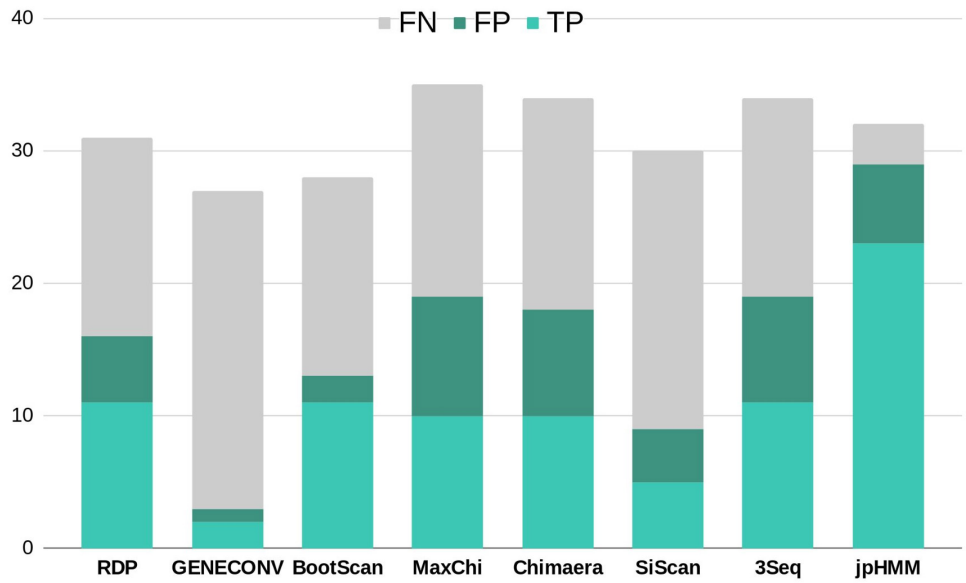


FIGURE 2 | Summary of true positive (TP), false positive (FP) and false negative (FN) recombination events detected by the different methods used in this study according to the results of ELW tests. RDP, GENECONV, BootScan, Maxchi, Chimaera, SiScan, and 3Seq were analyzed with RDP4.

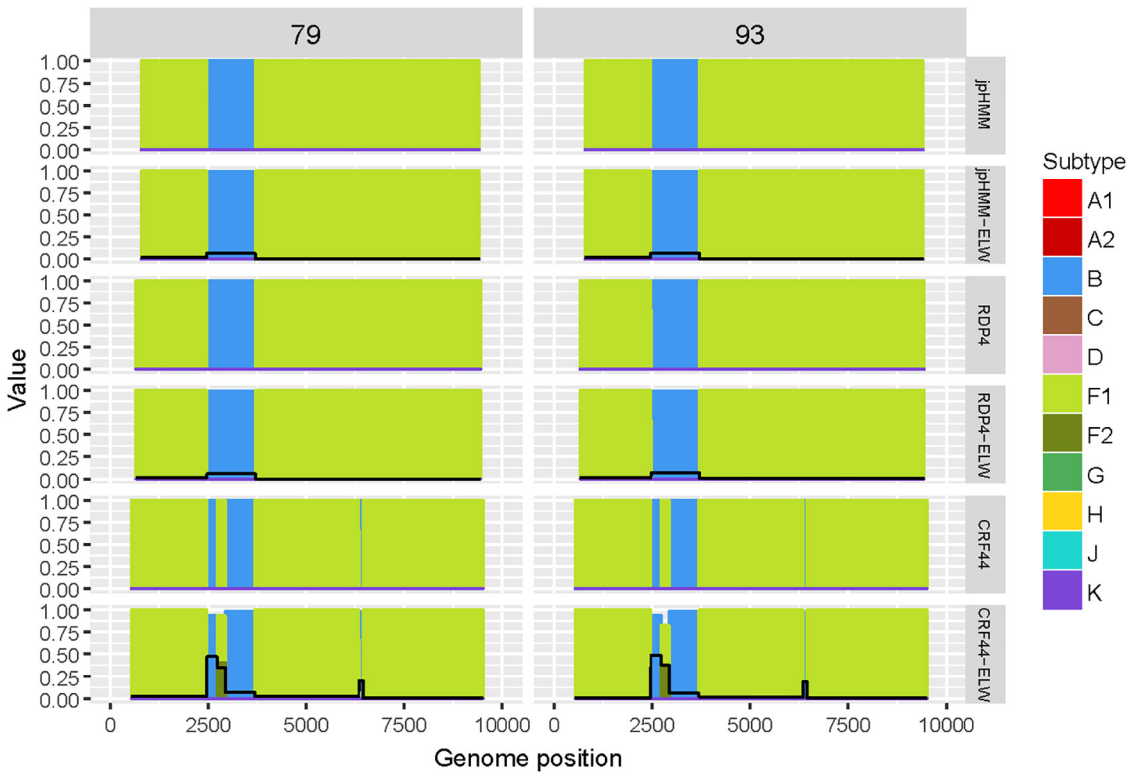


FIGURE 3 | Recombination analyses of samples 79 and 93 with recombination BPs determined by different sources (jpHMM, RDP4 and CRF44_BF reference BPs) and phylogenetic congruence (ELW) tests. The black line represents the proportion of unresolved quartets, which correspond to the inverse of the phylogenetic signal.

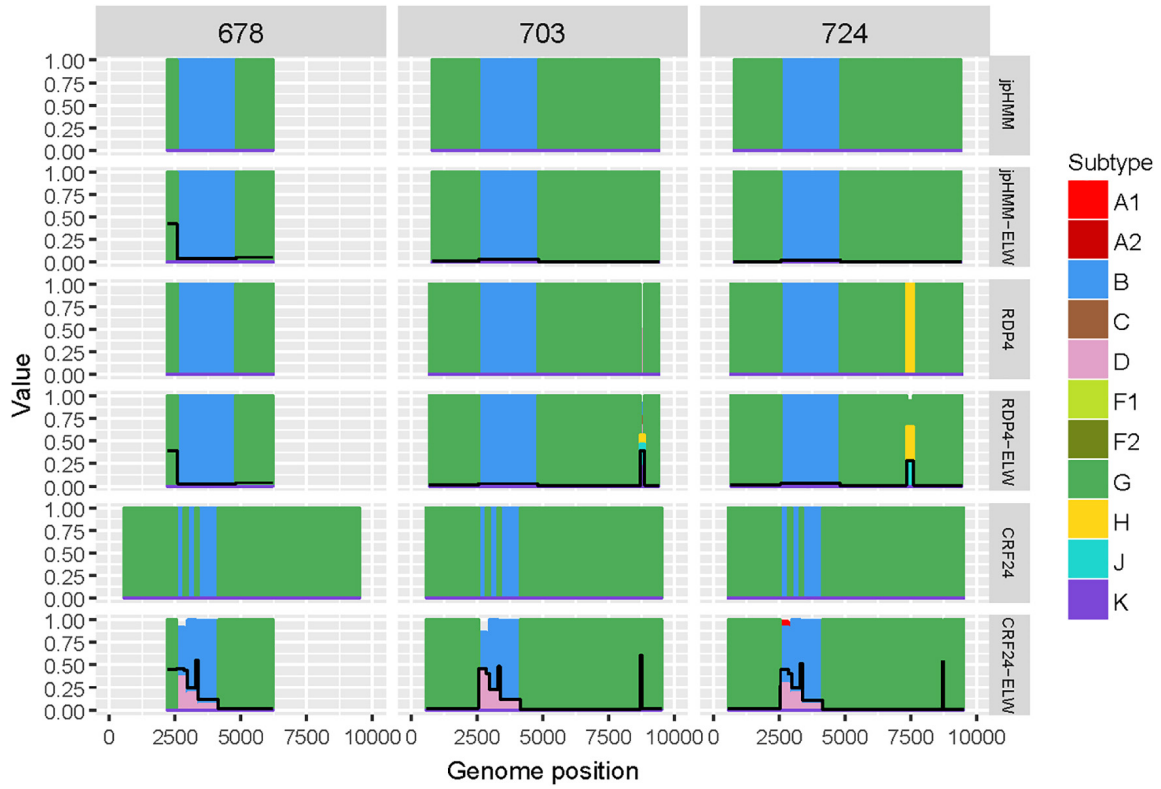


FIGURE 4 | Recombination analyses of samples 678, 703, and 724 with recombination BPs determined by different sources (jpHMM, RDP4, and CRF24_BG reference BPs) and phylogenetic congruence (ELW) tests. The black line represents the proportion of unresolved quartets, which correspond to the inverse of the phylogenetic signal.

in CRF24_BG, but recognized only the single B event detected jointly by RDP4 and jpHMM (**Figure 4**). It is important to note that in the case of sequence 678, the comparisons were limited because only a partial genome was obtained.

The two AD recombinants corresponded to samples 2011 and 2014 with A subtype being the major parental and with four and three fragments of subtype D, respectively. Initially, more recombination events of other subtypes were identified by jpHMM and RDP4. However, they were discarded based on the results of the ELW congruence tests. RDP4 failed to detect three events that were detected by jpHMM and further verified by tree-topology and phylogenetic signal (**Figure 5**). CRs analyses showed that the A portion corresponded to sub-subtype A3 whereas the D fragments clustered with D 'pure' or CRF19_cpx sequences (**Table 3**). Despite small discrepancies, 2011 and 2014 present a similar recombination pattern but, to our knowledge, no similar pattern has been reported to the databases yet.

The genome sequences derived from samples 3011 and 3164 were identified as complex forms, composed by more than two 'pure' subtypes: A, G, H, J, and K for 3011 and A, H, J, and K for 3164. For both samples, the A subtype was considered as the major parental although, in this case, there is little difference between the portion of the genome covered by the different subtypes. As before, the description of 3164 is limited because about only half of its genome sequence was obtained. These recombinant forms were very complex both in the detection and verification results (**Figure 6**). A total of 16 and seven different recombination events were detected for samples 3011 and 3164, respectively. JpHMM detected 10 events in sample 3011, which were different from the five events detected by RDP4. Only one event, 3011_7(K), was detected simultaneously by both methods (**Table 3**). None of the seven events detected in sample 3164, five by jpHMM and two by RDP4, was shared by the two methods.

Most of the events detected by jpHMM in sample 3011 (7/11) showed a low phylogenetic signal, with more than 25% of unresolved quartets (**Table 3**) whereas only one event detected by RDP4 in this sample failed this test. A similar result was observed for sample 3164, with three of the five events detected by jpHMM presenting low phylogenetic signal and with one of the two events detected by RDP4 also failing this test. This low signal might also affect subtyping inferences, with several discrepancies between the subtypes identified by jpHMM or RDP4 and those inferred as most likely by the phylogenetic congruence analyses (**Table 3**). Finally, the results of CRs analyses presented several cases of unclear relatedness. For instance, fragments of the major subtype of these two samples grouped within clade A but also with distinct URFs and complex forms, such as CRF45_cpx. The G and J fragments grouped with other complex forms, such as CRF13_cpx and CRF27_cpx. Finally, recombinant fragments

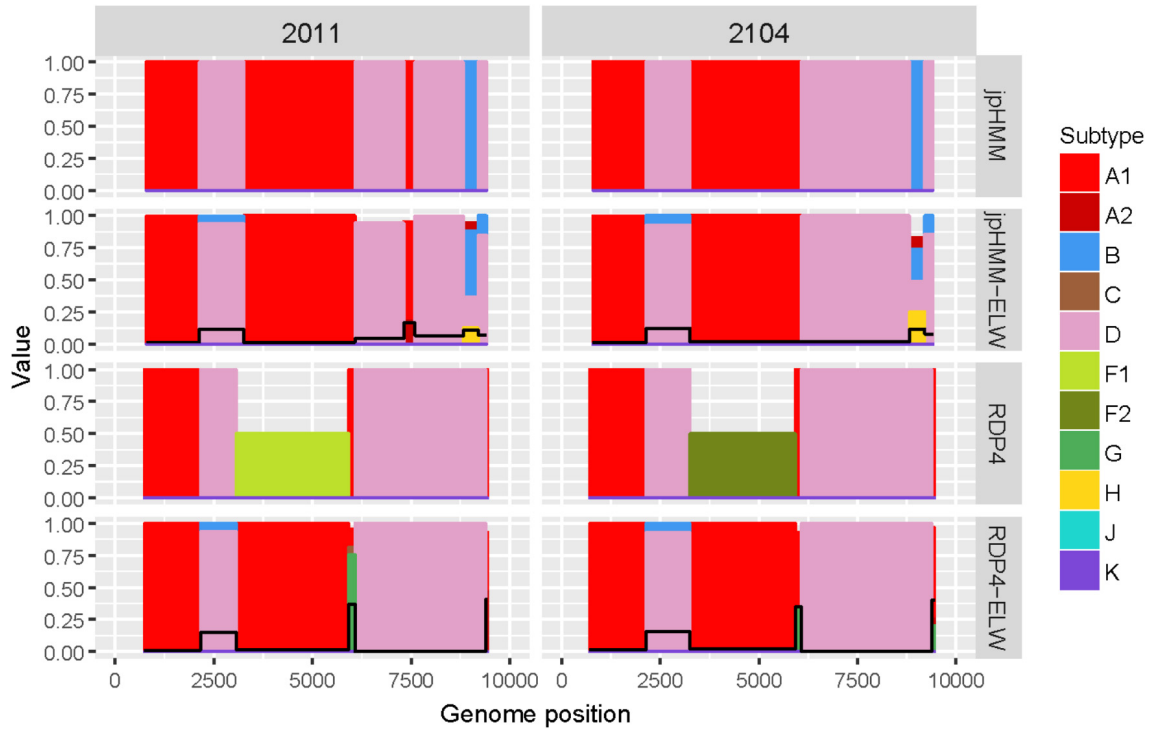


FIGURE 5 | Recombination analyses of samples 2011 and 2104 with recombination BPs determined by different sources (jpHMM and RDP4) and phylogenetic congruence (ELW) tests. Values for inconclusive results by RDP4 are arbitrarily represented as 0.5. The black line represents the proportion of unresolved quartets, which correspond to the inverse of the phylogenetic signal.

of subtype H clustered with 'pure' H sequences or CRF04_cpx and those of subtype K were related to K 'pure' sequences (**Table 3**). Altogether, the mosaic structure of these genomes could not be determined unambiguously. Samples 3011 and 3164 showed important discrepancies between them and at present, no CRFs_cpx with these patterns has been reported to the databases.

DISCUSSION

In this study, we have described seven nearly-complete and two partial genomes of different HIV-1 recombinant forms, detected in infected individuals from Comunitat Valenciana (Spain). We have performed detailed analyses to characterize their recombination events. Firstly, we identified putative recombinant fragments using different methods, jpHMM and seven of those included in RDP4. In a second phase, we corroborated the recombinant fragments by evaluating their phylogenetic signal, testing their topological congruence and analyzing their closest relatives using a comprehensive dataset of HIV-1 sequences including representatives of all the subtypes, CRF and URFs.

Two samples, 79 and 93, were initially identified as CRF38_BF based on the results of jpHMM and RDP4. However, our additional analyses of closest relatives and ELW congruence test revealed that both were related to CRF44_BF, a recombinant form first described in Chile in 2010 (Delgado et al., 2010), but already circulating in Madrid (Spain) in 2005 to 2007 (Gonzalez-Alba et al., 2011). These samples were obtained in 2012 and 2013, thus indicating that this recombinant form was actively circulating in Spain a few years later. BF recombinants might be a larger family than previously thought with differences in the location of breakpoints or length of the inserted fragments (Cevallos et al., 2017). This hypothesis is confirmed in this study, providing two additional BF genomes that were not properly characterized based on the breakpoints reported by two detection methodologies but identified with additional analyses. Subsequent mutations or differences in methodology might prevent the detection of the actual mosaic structure of additional recombinant forms in this family. Thorough phylogenetic analyses, including congruence tests, should always be performed to reliably infer the recombination structure of HIV-1 samples.

The three BG recombinants shared a similar mosaic structure, which was further verified in our subsequent analyses. They might constitute a SGR resulting from a new recombination event between CRF24_BG and subtype B. CRF24_BG was firstly identified in Cuba in 2007 (Sierra et al., 2007; Cevallos et al., 2017). However, it was found in several Spanish individuals in Madrid a few years ago (Gonzalez-Alba et al., 2011). This, along with the high prevalence of subtype B and the origin of two other BG recombinant forms, CRF14_BG and CRF73_BG (Delgado et al., 2002; Fernández-García et al., 2016), in the Iberian Peninsula, suggests a scenario that fits well with the

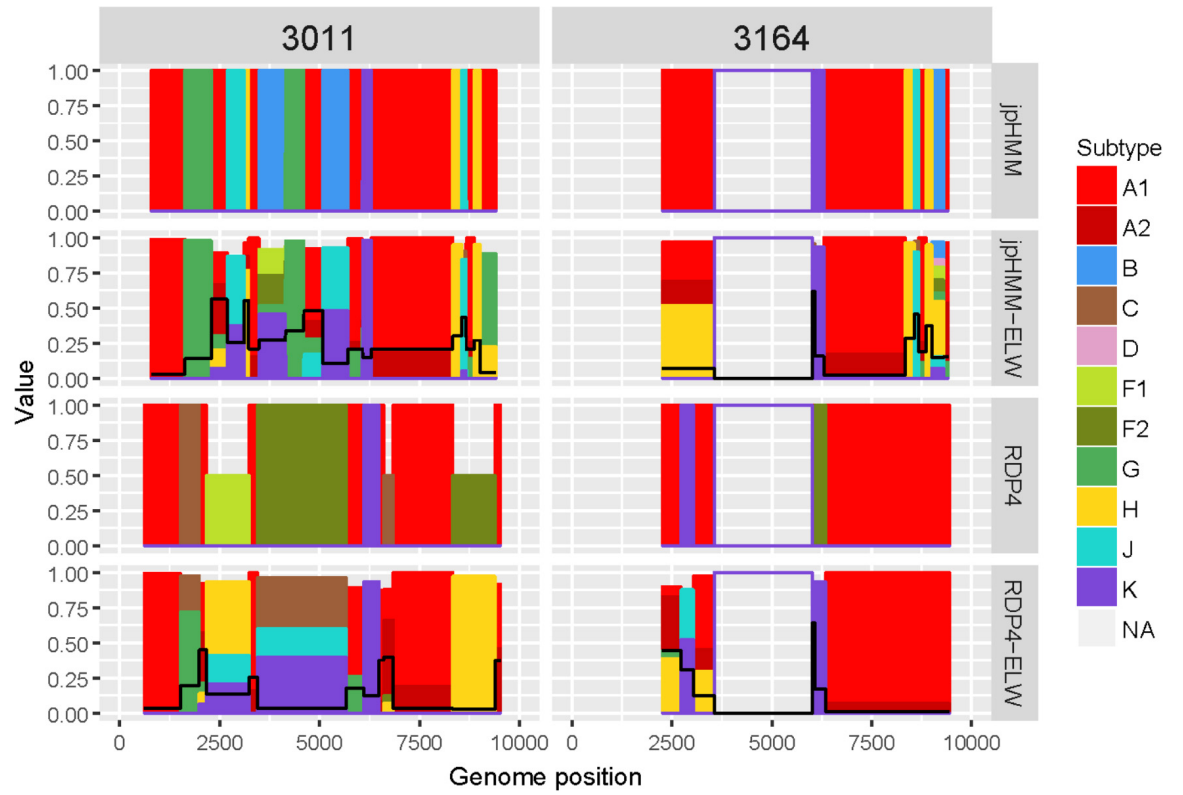


FIGURE 6 | Recombination analyses of samples 3011 and 3164 with recombination BPs determined by different sources (jpHMM and RDP4) and phylogenetic congruence (ELW) tests. Values for inconclusive results by RDP4 are arbitrarily represented as 0.5. The black line represents the proportion of unresolved quartets, which correspond to the inverse of the phylogenetic signal.

previous hypothesis as well as with the subsequent formation of new RFs with these subtypes.

Samples 2011 and 2014 might represent a new URF_AD, originated from several recombination events of 'pure' A3 and D sequences. In Spain, A and D subtypes are not frequent as they represent less than 1% of total infections (García et al., 2011; Gonzalez-Alba et al., 2011; Yebra et al., 2012; Patiño-Galindo et al., 2017) and no CRF_AD has been reported in this country so far. Hence, it is likely that this URF_AD has emerged in an endemic area for these subtypes and arrived in Spain by immigration.

Lastly, samples 3011 and 3164 represent complex forms composed of A, G, K, H and J subtypes that could not be identified accurately. Most of these subtypes, except G and A, are extremely rare in Spain (García et al., 2011; Gonzalez-Alba et al., 2011; Yebra et al., 2012; Patiño-Galindo et al., 2017). More specifically, in the Comunitat Valenciana, a previous analysis of the partial *pol* sequences of the 1,800 samples at the origin of this study revealed an important number of non-B infections (16.2%) but they did not include subtypes K, H, or J (Patiño-Galindo et al., 2017). These new URFs are an example of the enormous complexity of the recombination process in HIV-1 as several CRF_cpx are involved. The presence of CRFs from West-Central Africa, such as CRF9_cpx, CRF13_cpx, CRF27_cpx or CRF45_cpx, previously reported in Spain (Niama et al., 2009), might reflect an increased prevalence and generation of subsequent unique mosaic viruses in diverse geographical areas.

There are substantial differences in the performance of the eight methods used to detect recombination (**Figure 2**). The combined use of jpHMM and some (in our case at least three) methods implemented in RDP4 improves the reliable detection of recombination events: all the common events were further corroborated by ELW and phylogenetic signal analyses. This result agrees with previous suggestions of using several, methodologically different approaches to detect recombination events in HIV-1 (Posada, 2002). For those cases in which an event is not detected by one methodology, but it is positive according to others, additional corroboration analyses should be performed. In this case, 18 events detected by only jpHMM or RDP4 were later corroborated by ELW and phylogenetic signal analyses.

In general, RDP4 and jpHMM performed well for samples with one single recombination event. However, more false positive and negative events as well as cases of subtype misidentification were found for both programs with increasing complexity of the recombinant structures. Both methods showed the largest number of discrepancies in the detection of complex forms. The difficulties for the precise identification of the origin of these fragments were also evidenced by unresolved results in the congruence tests. Because some of these fragments are

very short, they may not have enough phylogenetic signal to differentiate between subtypes. This poses problems for the two methods used to detect them. In addition, the paucity of complete sequence information about 'rare' subtypes makes the detection and verification of recombinants, especially those involving complex forms, a difficult task.

CONCLUSION

We have identified nine recombinant forms circulating in the Comunitat Valenciana (Spain). Two of them are related to CRF44_BF but the rest correspond to new URFs. Some samples of this study revealed a complex recombination nature and origin. This might suggest that, apart from the endemic areas that exhibit the highest levels of genetic diversity and therefore represent recombination hotspots, other regions could be sources of recombinant progeny due to the increasing global gene flow (Foster et al., 2014). To better understand the viral diversity and dynamics, as well as the real impact of recombination on them, more efforts should be devoted to obtaining complete genome sequences and to perform in-depth analyses of recombination.

The present study highlights the utility of surveillance analyses of HIV-1 for additional goals to identify resistance mutations, such as studying genetic diversity and characterizing new recombinant forms. A thorough analysis of partial genome sequences obtained in the surveillance of antiretroviral resistance mutations may provide the best starting point for the discovery and characterization of new recombinant forms in HIV-1.

ETHICS STATEMENT

Approval of surveillance studies by an ethics committee is not required as per local legislation and host institution (FISABIO) guidelines.

AUTHOR CONTRIBUTIONS

The idea for this study was conceived by BB, MAB, and FG-C, who was also in charge of supervising the whole project. Data collection and reference searching were done by BB and MAB. The data analyses and drafting the manuscript were done by BB. The final manuscript was written, curated, and confirmed by BB, MAB and FG-C.

FUNDING

This research was supported by projects BFU2014-58656R and BFU2017-89594R from MINECO and MICU (Spanish Government), PROMETEO2016-0122 (from Generalitat Valenciana) and Conselleria de Sanitat Universal i Salut Pública (Generalitat Valenciana). BB was the recipient of an FPU contract for early stage researchers from MEC (Spanish Government).

REFERENCES

Abram, M. E., Ferris, A. L., Shao, G., Alvord, W. G., and Hughes, S. H. (2010). Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J. Virol.* 84, 9864–9878. doi: 10.1128/JVI.00915-10

Bonfield, J. K., Smith, K. F., and Staden, R. (1995). A new DNA sequence assembly program. *Nucleic Acids Res.* 23, 4992–4999. doi: 10.1093/nar/23.24.4992

Boni, M. F., Posada, D., and Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035–1047. doi: 10.1534/genetics.106.068874

Casado, G., Thomson, M. M., Delgado, E., Sierra, M., Vázquez-De Parga, E., Pérez-Alvarez, L., et al. (2003). Near full-length genome characterization of an HIV type 1 CRF05_DF virus from Spain. *AIDS Res. Hum. Retroviruses* 19, 719–725. doi: 10.1089/088922203322280955

Cevallos, C. G., Jones, L. R., Pando, M. A., Carr, J. K., Avila, M. M., and Quarleri, J. (2017). Genomic characterization and molecular evolution analysis of subtype b and bf recombinant HIV-1 strains among Argentinean men who have sex with men reveal a complex scenario. *PLoS One* 12:e0189705. doi: 10.1371/journal.pone.0189705

Cuevas, M. T., Fernández-García, A., Pinilla, M., García-Alvarez, V., Thomson, M., Delgado, E., et al. (2010). Short communication: biological and genetic characterization of HIV type 1 subtype B and nonsubtype B transmitted viruses: usefulness for vaccine candidate assessment. *AIDS Res. Hum. Retroviruses* 26, 1019–1025. doi: 10.1089/aid.2010.0018

Delgado, E., Ríos, M., Fernández, J., Pérez-Alvarez, L., Nájera, R., and Thomson, M. M. (2010). Identification of a new HIV type 1 BF intersubtype circulating recombinant form (CRF44_BF) in Chile. *AIDS Res. Hum. Retroviruses* 26, 821–826. doi: 10.1089/aid.2010.0006

Delgado, E., Thomson, M. M., Villahermosa, M. L., Sierra, M., Ocampo, A., Miralles, C., et al. (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* 29, 536–543. doi: 10.1097/00042560-200204150-00016

Fernández-García, A., Delgado, E., Cuevas, M. T., Vega, Y., Montero, V., Sánchez, M., et al. (2016). Identification of an HIV-1 BG intersubtype recombinant form (CRF73_BG), partially related to CRF14_BG, which is circulating in Portugal and Spain. *PLoS One* 11:e0148549. doi: 10.1371/journal.pone.0148549

Fernández-García, A., Pérez-Alvarez, L., Cuevas, M. T., Delgado, E., Muñoz-Nieto, M., Cilla, G., et al. (2010). Identification of a new HIV type 1 circulating BF intersubtype recombinant form (CRF47_BF) in Spain. *AIDS Res. Hum. Retroviruses* 26, 827–832. doi: 10.1089/aid.2009.0311

Foster, G. M., Ambrose, J. C., Hué, S., Delpech, V. C., Fearnhill, E., Abecasis, A. B., et al. (2014). Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of circulating recombinant form (CRF) 50_A1D. *PLoS One* 9:e83337. doi: 10.1371/journal.pone.0083337

García, F., Pérez-Cachafeiro, S., Guillot, V., Alvarez, M., Pérez-Romero, P., Pérez-Elías, M. J., et al. (2011). Transmission of HIV drug resistance and non-B subtype distribution in the Spanish cohort of antiretroviral treatment naïve HIV-infected individuals (CoRIS). *Antiviral Res.* 91, 150–153. doi: 10.1016/j.antiviral.2011.05.010

Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582. doi: 10.1093/bioinformatics/16.7.573

Gonzalez-Alba, J. M., Holguin, A., Garcia, R., Garcia-Bujalance, S., Alonso, R., Suarez, A., et al. (2011). Molecular surveillance of HIV-1 in Madrid, Spain: a phylogeographic analysis. *J. Virol.* 85, 10755–10763. doi: 10.1128/JVI.00454-11

Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2019). Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9

Holguin, A., Mulder, M., Yebra, G., Lopez, M., and Soriano, V. (2008). Increase of Non-B subtypes and recombinants among newly diagnosed HIV-1 native Spaniards and immigrants in Spain. *Curr. HIV Res.* 6, 327–334. doi: 10.2174/157016208785132455

Jia, L., Li, L., Li, H., Liu, S., Wang, X., Bao, Z., et al. (2014). Recombination pattern reanalysis of some HIV-1 circulating recombination forms suggest the necessity and difficulty of revision. *PLoS One* 9:e107349. doi: 10.1371/journal.pone.0107349

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Meth.* 14, 587–589. doi: 10.1038/nmeth.4285

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Lemey, P., Salemi, M., and Vandamme, A.-M. (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing.* Cambridge: Cambridge University Press.

Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563. doi: 10.1093/bioinformatics/16.6.562

Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., and Varsani, A. (2011). Recombination in eukaryotic single stranded DNA viruses. *Viruses* 3, 1699–1738. doi: 10.3390/v3091699

Martin, D. P., Posada, D., Crandall, K. A., and Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retroviruses* 21, 98–102. doi: 10.1089/aid.2005.21.98

Nadai, Y., Eyzaguirre, L. M., Constantine, N. T., Sill, A. M., Cleghorn, F., Blattner, W. A., et al. (2008). Protocol for nearly full-length sequencing of HIV-1 RNA from plasma. *PLoS One* 3:e1420. doi: 10.1371/journal.pone.0001420

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Niama, F. R., Vidal, N., Bazepeo, S. E., Mpoudi, E., Toure-Kane, C., Parra, H. J., et al. (2009). CRF45_AKU, a circulating recombinant from Central Africa, is probably the common ancestor of HIV Type 1 MAL and HIV type 1 NOGIL. *AIDS Res. Hum. Retroviruses* 25, 1345–1353. doi: 10.1089/aid.2009.0169

Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new Geminiviruses by frequent recombination. *Virology* 265, 218–225. doi: 10.1006/viro.1999.0056

Paradis, E., and Schliep, K. (2018). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Patiño-Galindo, J. Á, Torres-Puente, M., Bracho, M. A., Alastrué, I., Juan, A., Navarro, D., et al. (2017). The molecular epidemiology of HIV-1 in the Comunidad Valenciana (Spain): analysis of transmission clusters. *Sci. Rep.* 7:11584. doi: 10.1038/s41598-017-10286-1

Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., and González-Candelas, F. (2015). Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect. Genet. Evol.* 30, 296–307. doi: 10.1016/j.meegid.2014.12.022

Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* 19, 708–717. doi: 10.1093/oxfordjournals.molbev.a004129

Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13757–13762. doi: 10.1073/pnas.241370698

Revell, L. J. (2011). Phytools: an R package for phylogenetic comparative biology (and other things). *Meth. Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210x.2011.00169.x

Rodrigo, A. G., Shpaer, E. G., Delwart, E. L., Iversen, A. K. N., Gallo, M. V., Brojatsch, J., et al. (1999). Coalescent estimates of HIV-1 generation time *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2187–2191. doi: 10.1073/pnas.96.5.2187

Ruchansky, D., Casado, C., Russi, J. C., Arbiza, J. R., and Lopez-Galindez, C. (2009). Identification of a new HIV type 1 circulating recombinant form (CRF38_BF1) in Uruguay. *AIDS Res. Hum. Retroviruses* 25, 351–356. doi: 10.1089/aid.2008.0248

Schultz, A.-K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., et al. (2010). jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* 38, 1059–1059. doi: 10.1093/nar/gkp371

Shriner, D., Rodrigo, A. G., Nickle, D. C., and Mullins, J. I. (2004). Pervasive genomic recombination of HIV-1 *in vivo*. *Genetics* 167, 1573–1583. doi: 10.1534/genetics.103.023382

Sierra, M., Thomson, M. M., Posada, D., Pérez, L., Aragonés, C., González, Z., et al. (2007). Identification of 3 phylogenetically related HIV-1 BG intersubtype circulating recombinant forms in Cuba. *J. Acquir. Immune Defic. Syndr.* 45, 151–160. doi: 10.1097/qai.0b013e318046ea47

Sierra, M., Thomson, M. M., Ríos, M., Casado, G., Ojea-de Castro, R., Delgado, E., et al. (2005). The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. *Infect. Genet. Evol.* 5, 209–217. doi: 10.1016/j.meegid.2004.07.010

Simon-Loriere, E., and Holmes, E. C. (2011). Why do RNA viruses recombine? *Nat. Rev. Microbiol.* 9, 617–626. doi: 10.1038/nrmicro2614

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129.

Strimmer, K., and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. B Biol. Sci.* 269, 137–142. doi: 10.1098/rspb.2001.1862

Thomson, M. M., Delgado, E., Herrero, I., Villahermosa, M. L., Vázquez-de Parga, E., Cuevas, M. T., et al. (2002). Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J. Gen. Virol.* 83, 107–119. doi: 10.1099/0022-1317-83-1-107

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer.

Yebra, G., de Mulder, M., Martín, L., Rodríguez, C., Labarga, P., Viciana, I., et al. (2012). Most HIV type 1 non-Binfections in the Spanish cohort of antiretroviral treatment-naïve HIV-infected patients (CoRIS) are due to recombinant viruses. *J. Clin. Microbiol.* 50, 407–413. doi: 10.1128/JCM.05798-11

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 May 2019
doi: 10.3389/fmicb.2019.01096

Characteristics of Envelope Genes in a Chinese Chronically HIV-1 Infected Patient With Broadly Neutralizing Activity

*Dai Zhang[1,2], Sen Zou[1], Yuanyuan Hu[1], Jiali Hou[1], Xintao Hu[1†], Li Ren[1], Liying Ma[1], Xiang He[3], Yiming Shao[1] and Kunxue Hong[1]**

[1] State Key Laboratory of Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing, China, [2] The First Affiliated Hospital of Henan University of Traditional Chinese Medicine, Zhengzhou, China, [3] Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China

Exploring the characteristics of the HIV-1 envelope glycoprotein (*env*) gene in a natural HIV-1 infected individual, with broadly neutralizing activity, may provide insight into the generation of such broadly neutralizing antibodies and initiate the design of an appropriate immunogen. Recently, a chronically HIV-1 infected patient with broadly neutralization activity was identified and a VRC01-class neutralizing antibody DRVIA7 (A7) was isolated from the patient. In the present study, 155 full length HIV-1 *env* gene fragments (including 68 functionally Env clones) were amplified longitudinally from the plasma of six time points spanning over 5 years in this donor. Viral features were analyzed by comparing Env clones of different time points, as well as 165 Chinese HIV-1 subtype B *env* sequences from HIV Sequence Database (Chinese B_database). Shorter V1 length, less potential glycan and a lower ratio of NXT: NXS in gp160 were observed in the first five time points compared to that from the last time points, as well that from the Chinese B_database. A sequence analysis and a neutralization assay of Env-pseudoviruses showed that the increasing diversity of *env* sequences in the patient was consistent with the appearance and maturation of A7 lineage antibodies. The potent neutralization activity and viruses that escaped from the neutralization of the concurrent autologous plasma, are consistent with higher residue variations at the antibody recognition sites. Almost all viruses from the plasmas were neutralization-resistant to VRC01 and A7 lineage antibodies. For a chronically HIV-1 infected individual over 10 years, we found that greater viral diversity, short V1 sequences and less potential N-linked glycosylation (PNGS) in V1, might be associated with the development of broadly neutralizing antibody responses.

Keywords: HIV-1, envelope gene, broadly neutralizing antibodies, glycosylation, diversity

OPEN ACCESS

Edited by:
Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:
Yuh Koon Tong,
Sunway University, Malaysia
Bernard A. P. Lafont,
National Institute of Allergy and Infectious Diseases (NIAID), United States

***Correspondence:**
Kunxue Hong
hongkx@chinaaids.cn

†Present address:
Xintao Hu,
Human Retrovirus Pathogenesis Section, Vaccine Branch, Center for Cancer Research, National Cancer Institute, Frederick, MD, United States

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *29 November 2018*
Accepted: *30 April 2019*
Published: *24 May 2019*

Citation:
Zhang D, Zou S, Hu Y, Hou J, Hu X, Ren L, Ma L, He X, Shao Y and Hong K (2019) Characteristics of Envelope Genes in a Chinese Chronically HIV-1 Infected Patient With Broadly Neutralizing Activity. Front. Microbiol. 10:1096. doi: 10.3389/fmicb.2019.01096

INTRODUCTION

HIV-1 was identified as the pathogen of acquired immunodeficiency syndrome (AIDS) three decades ago (Gottlieb and Schroff, 1981; Barre-Sinoussi et al., 1983; Gallo et al., 1983; Dalgleish et al., 1984); however, the development of efficient and safe vaccines is still under way (Moore, 2018; Burton, 2019). The extreme virus diversity contributes to the main challenge for HIV-1 vaccines development. HIV-1 contains four groups: M, O, N, and P, and the dominant group M is further subdivided into nine distinct subtypes and increasing circulating recombinant forms (Robertson et al., 2000; Richman et al., 2003; Hemelaar, 2012). Broadly neutralization antibodies (bNAbs) targeting Env can protect animal models from the challenge of SHIV-1, neutralize most global circulating strains, and accelerate elimination of HIV-1-infected cells (Lu et al., 2016; Julg et al., 2017), therefore, eliciting bNAbs is an important goal of HIV-1 vaccines. However, in numerous pre-clinical and clinical trials of HIV-1 vaccines, bNAbs have not been successfully induced thus far (Burton and Hangartner, 2016; Kwong and Mascola, 2018).

Despite immune strategies to prevent HIV-1 infection have not been discovered, a recent study showed that broadly neutralizing activity can be detected in about 50% of HIV-1 infected individuals (Hraber et al., 2014), indicating that the human immune system indeed has the ability to elicit a bNAbs response. Before 2009, only four bNAbs were available; b12, 2F5, 4E10, and 2G12 (Muster et al., 1993; Buchacher et al., 1994; Conley et al., 1994; Trkola et al., 1996). Recently, with the wide utilization of new technologies such as single cell antibody cloning techniques, micro-neutralization assay, and B cell repertoire analysis (Simek et al., 2009; Walker et al., 2009, 2011; Huang et al., 2013; Zhu et al., 2013), hundreds of bNAbs were successively isolated from HIV-1-infected individuals (Burton and Hangartner, 2016; Wu and Kong, 2016). BNabs isolated in the natural HIV-1 infection provide a prototype that could be elicited by vaccines (Van Gils and Sanders, 2013; Bonsignori et al., 2017; Haynes and Mascola, 2017). A few studies tracing the evolution of bNAbs from the time of HIV-1 infection have revealed that viral and antibody evolution led to the induction and maturation of the bNAbs lineage (Liao et al., 2013; Gao et al., 2014; Bonsignori et al., 2016). However, given the huge diversity of HIV-1 and high complexity of the interaction between HIV-1 and the immune system, the development pathway of bNAbs was not all identical, even for a class of bNAbs in different individuals (Zhou et al., 2015). Thus, further exploring the general characteristics underlying the development of bNAbs would provide insights into efficient vaccines.

In our previous study, a long-term non-progressor (LTNP) DRVI01 with broadly neutralization activity was identified (Hu et al., 2012), and DRVIA7(A7), a VRC01-like broadly monoclonal neutralizing antibody targeting CD4 binding site (CD4bs) was isolated from this patient (Kong et al., 2016). Systematic analysis of the development of A7 over 5 years showed that the heavy chain of the antibody rapidly matured within 2 years, while the barrier of glycans of the gp120 protein blocked the development of the light chain of the antibody. However, the viral Env characteristics of DRVI01 have not been elaborated in detail.

In the present study, 155 full length HIV-1 *env* gene fragments were amplified longitudinally from the DRVI01 plasma of six time points spanning 5 years. Viral features were analyzed by comparing Env clones of different time points, as well as 165 Chinese HIV-1 subtype B Env sequences from the HIV Sequence Database (Chinese B_database). Sixty-eight functional Env clones were expressed as pseudoviruses to test neutralization sensitivities to autologous plasmas, representative bNAbs and A7 lineage reconstituted antibodies, respectively. The mutations of critical residues in the contact region of VRC01 were also analyzed. The results showed that for a chronically HIV-1 infected individual over 10 years, the greater viral diversity, short V1 sequences and less potential N-linked glycosylation (PNGS) in V1, might be associated with the development of broadly neutralizing antibody responses.

MATERIALS AND METHODS

Study Subject

The samples described in this study were collected from an HIV-1-infected Chinese patient, DRVI01 who became HIV-1 infected by clade-B' strain during commercial plasma donation between 1992 and 1995 (Hu et al., 2012; Kong et al., 2016). We collected the blood sample of the patient every 6 months between 2005 and 2010. The patient was antiretroviral treatment (ART)-naive, and the range of viral load of six time points ranged from 74,200 to 310,000 copies/ml and the CD4+ T cell count ranged from 335 to 769 cells/μl (**Table 1**). DRVI01 was identified as a broadly cross-reactive neutralizer, whose plasma exceeded a 95% neutralizing breadth against a panel of 25 viruses at all six time points. From PBMC of the subject, five neutralizing antibodies (NAbs) with limited neutralization breadth (all <40% breadth), including a VRC01-like neutralizing antibody DRVIA7, were isolated (Kong et al., 2016). The study was reviewed and approved by the Institutional Review Board of the National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention. The subject provided written informed consent before blood and data collection.

Viral RNA Extraction, cDNA Synthesis, and Single-Genome Amplification

Viral RNA was extracted from the plasma using a QIAamp viral RNA mini kit (Qiagen, Valencia, CA) and subjected to first-strand cDNA synthesis immediately using the SuperScript III reverse transcriptase (Invitrogen Life Technologies, Grand Island, NY). Single-genome amplification (SGA) of the full-length gp160 gene was performed as described previously (Wu et al., 2012). Briefly, the synthesized cDNA was serially diluted and distributed in replicates of 12–16 PCRs in Thermo Grid 96-well plates, to identify a dilution where PCR-positive wells constituted about 30% of the total number of reactions. The SGA criteria of fewer than 30% positive results was acquired, and most of the wells contained amplicons derived from a single cDNA molecule in the suitable dilution.

TABLE 1 | Characteristics of DRVI01 donor

Sample date	CD4+ T cells (cell/μl)	Viral load (copies/ml)	Mean *env* distance ± SD (%)	No. of *env* sequences	No. of functional Env clones
2005-7-12	475	1.27E+05	3.06 ± 1.35	29	14
2005-10-18	438	2.28E+04	3.17 ± 1.21	25	12
2006-4-6	335	7.42E+04	3.48 ± 1.63	34	15
2008-3-20	747	3.29E+05	4.79 ± 2.04[a]	32	18
2009-5-12	512	2.68E+05	5.26 ± 2.22[b]	14	N/A
2009-12-25	769	3.10E+05	5.36 ± 2.36[c]	21	9

[a,b,c]*Compared with 2006-4-6, 2005-10-18, and 2005-7-12, all $P < 0.05$.*

DNA Sequencing, Alignment, and Phylogenetic Analyses

SGA products were sequenced on an ABI 3770 Sequencer (Applied Biosciences). The full-length gp160 gene fragments for each amplicon were assembled and edited using Sequencher 4.1 (Gene Codes, Ann Arbor, MI). All chromatograms were inspected for sites of mixed bases (double peaks), which would provide evidence of priming from more than one template or the introduction of a PCR error in early cycles. Any sequence with evidence of double peaks was excluded from further analysis.

Phylogenetic and evolutionary analysis was conducted using MEGA 6. The *env* sequences were aligned using Gene Cutter.[1] The nucleotide sequences together with B.CN.RL42.U71182, a Chinese B' reference, were initially aligned and then checked by hand using BioEdit. The protein phylogenetic tree was built by the Neighbor-Joining method with the Jones–Taylor–Thornton model. After gap striping, the nucleotide phylogenetic tree was reconstructed by the maximum-likelihood method with GTR+Γ4+I substitution model. The reliability of internal nodes was assessed by a bootstrap test (1000 replicates). Genetic diversity of the Env variants from all time points was indicated as mean gene distances, which were calculated by MEGA 6.0 with the Bootstrap method and Kimura 2-parameter model.

Variable Region Length and Gp160 PNGS Analyses

After the amino acid sequence alignment, the variable region length and the number of PNGS were determined using the online tool Variable Region Characteristics for V1, V2, V3, V4, V5.[2] For comparison, a set of pre-aligned 165 HIV-1 subtype B Env protein sequences from China (Chinese B_database) were downloaded from the Los Alamos HIV database[3] on June 15, 2017. The criteria for this data set were subtype B, intact gp160 sequences, Chinese, and one sequence per donor. The comparison of length variation and the V1 (aa 131–149), V2 (aa 158–197), V3 (aa 296–331), V4 (aa 385–418), and V5 (aa 460–471) loops in gp120 between different time points and the Chinese B_database were calculated by counting the number of amino acids. The number of PNGS and number of NXT or NXS motifs (X is any amino acid residue except proline) were identified using the N-Glycosite at the Los Alamos HIV database website.[4]

Antibodies Used in the Study

BmAbs PGT121, PGT135, 2G12, 10E8, 12A21, and VRC01 were received from NIH AIDS Research and Reference Reagent Program (Trkola et al., 1996; Zhou et al., 2010; Walker et al., 2011; Huang et al., 2012, 2016). The heavy chains of A7 reconstituted antibodies were derived from DRVIA7H variants of 2006 and 2009, the light chains of antibodies were from a VRC01 light chain and 2009 DRVIA7L repertoire (Kong et al., 2016). We selected four reconstituted A7 antibodies in addition to DRVIA7 with a neutralization breadth ranging from 24 to 88% and the characteristics of A7 antibodies referred to Kong et al. (2016).

Pseudovirus Preparation, Titration, and Neutralization Assays

Pseudoviruses were prepared, titrated as previously described (Li et al., 2005). Briefly, exponentially dividing 293T cells were cotransfected with Env/Rev expression plasmid and an Env-deficient HIV-1 backbone vector (pSG3△Env). Pseudovirus-containing supernatant was harvested 48 h post-transfection, and filtered (0.45 μm pore size) and single-use aliquots (1 ml) were stored at −80°C. The 50% tissue culture infectious dose (TCID50) of a single-thawed aliquot of each pseudovirus batch was determined in TZM-b1 cells.

Neutralization was measured as a reduction in Luc reporter gene expression after a single round of virus infection in TZM-bl cells, as described previously (Li et al., 2005). Briefly, 200 TCID50 of pseudovirus was incubated with serial threefold dilutions of plasma sample or antibodies for 1 h at 37°C. Freshly TZM-b1 cells were added. One set of control wells received cells only, and another set received cell plus pseudovirus. Following 48 h incubation, 150 μl culture was removed and 100 μl luciferase reporter gene assay system reagent (Bright-Glo; Promega) was added and incubated 2 min. 150 μl lysate from each well was transferred to 96-well black solid plates for measurement of luminescence using a luminometer (PerkinElmer Life Sciences). The 50% inhibitory dose (ID50) was defined as either the plasma dilution or sample concentration at which relative

[1]https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html

[2]https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/index.html

[3]https://www.hiv.lanl.gov/

[4]http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html

luminescence units (RLU) were reduced 50% compared to virus control wells.

Statistical Analyses

SPSS software was used in the data analyses. Differences in the number of amino acids and potential N-linked glycans in the Env protein were compared using a one-way ANOVA, and the independent-samples *T* test was used between groups. Differences were considered significant if $P < 0.05$. The two-sided Fisher's exact test was used to determine the difference of relative loss of specific PNGS between groups.

RESULTS

Phylogenetic Analyses of the GP160 Genes

Env protein is the major target of NAbs. HIV-1 can escape the neutralization of NAbs with substitution, insertion or deletion in Env. To examine the Env evolution in DRVI01, SGA was used to isolate *env* genes from plasma samples. Around 20 (14–34) intact *env* genes (about 10 functional) were derived from each time point (**Table 1**). The virus diversity increased gradually from 2005-7 to 2009-12, and compared with those of the first three time points, significantly higher *env* diversification were found in the later three time points (**Table 1**).

Phylogenetic analysis was performed to examine the relationship of these Env sequences, both protein sequence tree (**Figure 1A**) and nucleotide sequence tree (**Figure 1B**) displayed similar clustering mode. Most of the sequences formed distinct time-specific lineages, with a fraction of sequences intermingled. The phylogenetic trees, especially the nucleotide sequence one, showed that a main viral population and a minor viral population evolved in parallel in the patient for more than 4.5 years, and the latter was not eliminated over time. The gene distance of all time points for all sequences is 0.028876 ± 0.002026 (mean ± SE). The distances of minor branch and main branch were 0.01963 ± 0.00144, 0.02372 ± 0.00152, respectively. Compared with the minor population, the main viral population showed higher diversity, which may reflect the better adaptation of HIV-1 in response to host selective pressures (**Figure 1B**). However, no differences in neutralization sensitivity of viruses between the main and minor branches were observed (**Table 3**).

Significantly More Potential PNGS and the NXT:NXS Ratio of Gp160 at the Later Time Points

Env uses multiple mechanisms to escape from host immune response, including amino acid substitutions, insertions in the variable domains and increasing PNGS on its outer surface. In this study, we compared the amino acid length, PNGS and ratio of NXT:NXS at the six time points to observe the features of gp160. The results showed the length of gp160 was relatively constant (**Figure 2A**), however, the number of PNGS in gp160 were significantly more at 2009-12 time point than those at the first five time points (**Figure 2B**). It has been reported that NXS motifs have a two to three times lower probability of becoming glycosylated than NXT motifs (Kaplan et al., 1987; Gavel and Heijne, 1990). The actual extent of glycosylation of a given Env molecule may therefore not entirely equal to the total number of PNGS. In the study, we found the ratio of NXT: NXS in gp160 at the first four time points (07/2005–03/2008) were significantly lower than those at the last two time points (05/2009, 12/2009) and the Chinese B_database (all $P < 0.05$) (**Figure 2C**). Altogether, amino acid substitutions and increasing PNGS are likely the main ways that viruses adapt the host immune response in the patient.

Shorter V1 Length and Fewer PNGS in V1 at the First Five Time Points

To observe the features of the Env sequences from different time points, we further compared variable regions length and numbers of PNGS and the ratio of NXT:NXS from six time points as well as the Chinese B_database. The results showed that a shorter V1 region and fewer numbers of PNGS in V1 at the first five time points compared with those at 2009-12 time point and Chinese B_database (**Figure 3**). Heavily glycosylated V1V2 loop locates in the apex of Env spike were observed, and the loop can obstruct the exposure of co-receptor and CD4 binding sites (Wyatt et al., 1995; Rusert et al., 2011). In general, an increase of the length of V1V2 and the number of PNGS would help viruses escape autologous antibody neutralization and shield the more conserved domains associated with receptor binding (Van Gils et al., 2011; Wu et al., 2012). Our results were inconsistent with previous reports, as the length of theV1V2 loop was not significantly increased in the subject, after a HIV-1 infection over 10 years. Some studies considered that early HIV-1 variants with shorter variable V1V2 loop correlated with the development of later cross-reactive neutralizing activity (Rademeyer et al., 2007; van den Kerkhof et al., 2013). At the first four time points, the lower ratio of NXT:NXS indicated lower probability of glycosylation, combining with shorter V1 length and fewer PNGS may favor exposure of interior conservative epitopes on Env and the binding of the B cell receptor to interior epitopes.

Increasing Env Diversity Consistent With the Development of A7 Antibodies

Co-evolution of the virus-antibody in the HIV infection lead to the induction and development of bNAbs, and Env diversification in contact residues of bNAbs preceded the development of the neutralization breadth (Liao et al., 2013; Bonsignori et al., 2016). Therefore, the characteristics of evolutionary modes of contact residues of gp120 and VRC01 in HIV infection could provide valuable insight in designing the sequential immunogens. Loop D plays an important role in the interaction of VRC01 and gp120, and recurrent mutations in loop D were found in VRC01-resistant viruses (Zhou et al., 2010; Li et al., 2011). By longitudinally tracking the evolution of Env from six time points, we found that the dominant variants presented more mutations in loop D at a later time point (**Figure 4**). Interestingly, the highest diversification of loop D was found at the 2006-04 time point, and the time coincided with the emergence of

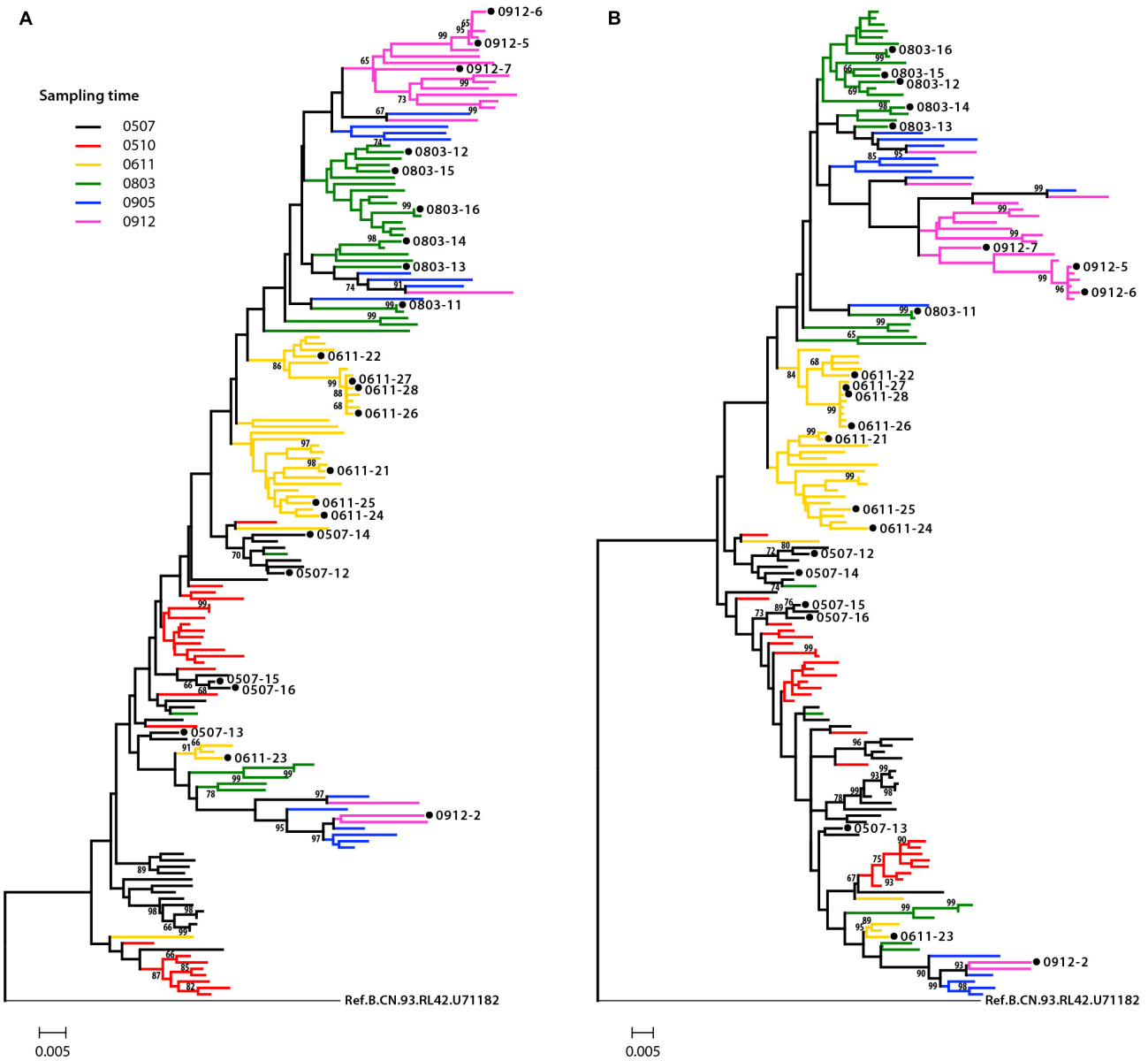


FIGURE 1 | Phylogenetic tree of HIV-1 envelope sequences. A total of 155 gp160 sequences from an HIV-1-infected donor DRVI01 were aligned with the reference sequence B.CN.RL42.U71182. The tree was constructed based on sequence distance and rooted at B.CN.RL42.U71182 for visualization. **(A)** The protein phylogenetic tree was built by the Neighbor-Joining method with the Jones–Taylor–Thornton model. **(B)** The nucleotide phylogenetic tree was reconstructed by the maximum-likelihood method with GTR+Γ4+I substitution model. The tree showed that a main viral population and a minor viral population have evolved in parallel in the patient for more than 4.5 years, and the latter is not eliminated over time. The gp160 sequences were color coded as follows: black, June 2005; red, October 2005; yellow, 2006; green, 2008; blue, May 2009; purple, December 2009. The horizontal branch scale is indicated for each tree.

the A7 lineage, suggesting that the diversification of loop D could be associated with the induction of the A7 lineage. The CD4 binding loop was one of the critical contact regions of VRC01 and gp120. Mutations in this region could result in the loss of the capability of virus infection. In our study, the CD4 binding loop exhibited lower diversification, spanning all six time points. The V5/β24 loop was another contact region of VRC01 and gp120. Variants in the region were observed in the VRC01 resistant isolates (Zhou et al., 2010; Li et al., 2011). In line with our observation in loop D, most of the mutations in V5/β24 were focused in the tip of the V5 loop, and more diversification in V5 were found in the dominant variants at a later time point (**Figure 4**).

Altogether, three main contact regions of VRC01 and gp120 presented multiple mutations at all six time points. Higher residue variation at antibody recognition sites indicated that the viral and antibody evolution might lead to induction and maturation of A7 lineage antibodies.

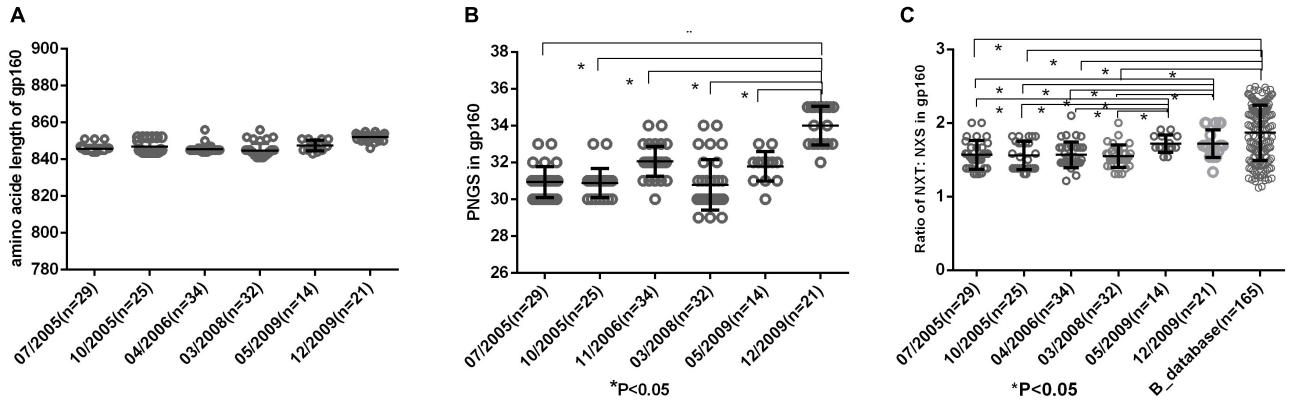


FIGURE 2 | (A) Comparison of amino acid length of gp160, which indicted no significant differences in amino acid length of gp160 among different time points. **(B)** Comparisons of glycan number of gp160 in six time points, which showed an increasing glycan number from 07/2005 to 12/2009 and the glycan number in gp160 from 12/2009 was significantly higher than the first five time points (07/2005–05/2009) (all $P < 0.05$). **(C)** Ratio of NXT:NXS of gp160 in six time points and Chinese B_database. NXT: NXS ratios in gp160 from the first four time points were significantly lower than those from two time points in 2009 and Chinese B_database (all $P < 0.05$).

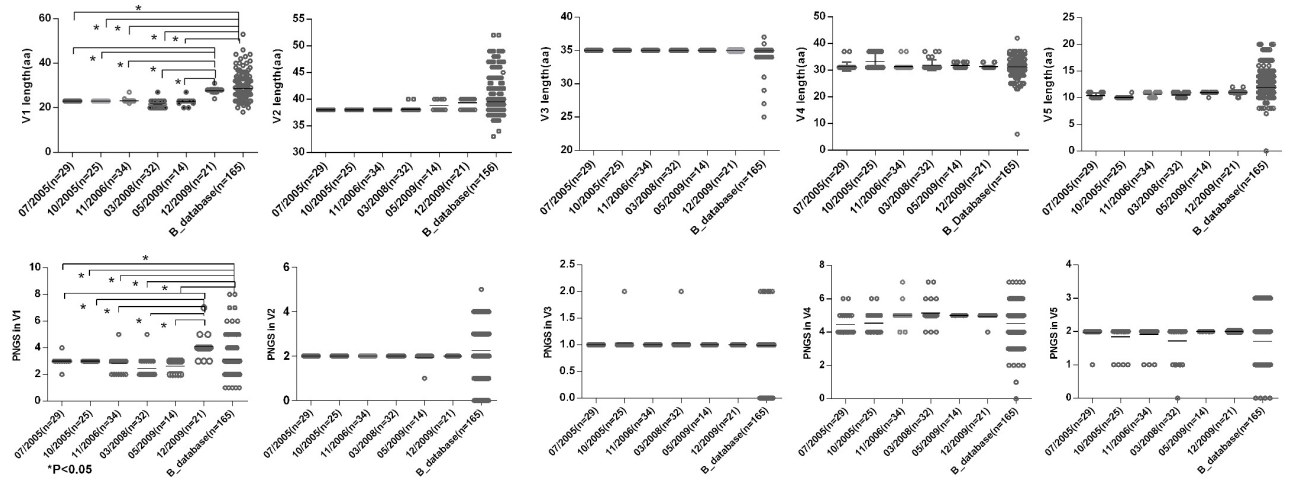


FIGURE 3 | Comparisons of sequence length and glycan number between DRVI01 and Chinese B-database. Compared with that of Chinese B_database, the V1 region lengths of time points 07/2005, 10/2005, 04/2006, 03/2008, 05/2009 were significantly shorter (all $P < 0.05$); the V1 length of 12/2009 was the longest among six time points. As well, the glycan numbers in V1 of Chinese B_database and12/2009 were more than that of five time points (all $P < 0.05$). The V2 length of two time points in 2009 and Chinese B_database were slightly increased than that of the first four time points.

Antibody-Based Selection Pressure Driving Ongoing Viral Evolution

Plasma from six time points of DRVI01 exhibited potent neutralization activity and can neutralize over 95% of viruses in a panel of 25 pseudoviruses (Kong et al., 2016). To observe the interaction between viruses and autologous plasmas, the neutralization sensitivity of the functional Env variants of five time points was measured against autologous plasma samples (**Table 2**). In line with a previous observation, autologous plasmas can potently neutralize viruses from earlier time points but cannot neutralize viruses from concurrent and later time points (Bunnik et al., 2008; Moore et al., 2009a,b; Wu et al., 2012). The results indicated that even a robust immune response can efficiently neutralize autologous viruses; HIV-1 managed to escape the neutralization of NAbs by mutations over time. Consistent with phylogenetic analysis, antibody-based selection pressure drove the ongoing viral evolution.

Neutralization Sensitivity of the Pseudoviruses Against the bNAbs Consistent With Variation in Critical Residues

The neutralization sensitivity of the Env pseudoviruses from different time points was tested against six bNAbs including PGT121, PGT135, 2G12, VRC01, 10E8, and12A21; and the mutations of critical residues in these bNAbs epitopes were analyzed for each Env variant (**Table 3** and

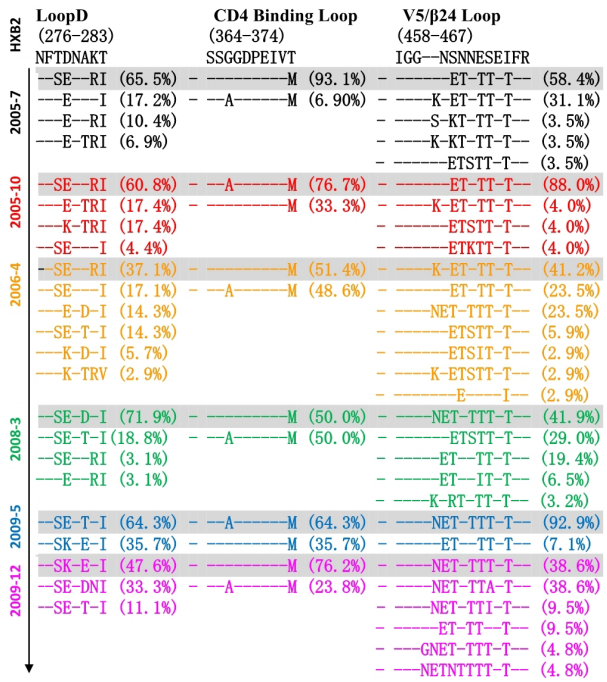


FIGURE 4 | Amino acid evolutionary modes located in the contact regions of gp120 and VRC01 including loop D (276–283), CD4 binding loop (364–374) and V5/β24 loop (458–467) in six time points (HXB2 numbering is indicated). Different amino acids sequences in three contact regions were sorted and denoted percentage in 155 Env sequences of six time points, and – (dashed) represents same residues with HXB2. The sequences were color coded as follows: black, June 2005; red, October 2005; yellow, 2006; green, 2008; blue, May 2009; purple, December 2009. The dominant variants presented more mutations in loop D at later time point, and the highest diversification of loop D was found at 2006-04 time point. CD4 binding loop exhibited lower diversification spanning all six time points, and more diversification in V5 was found in the dominant variants at later time point.

Supplementary Table 1). Almost all pseudoviruses were found to be resistant to VRC01 and 12A21, suggesting the presence of CD4bs antibodies pressure over the study period, which was consistent with the strong CD4bs antibodies specificity and isolation of a VRC01-class antibody A7 from the subject (Kong et al., 2016). Binding interface of VRC01 on gp120 largely located in Loop D, CD4 binding loop and V5/β24; viruses can abrogate VRC01-mediated neutralization by key residue mutations in these regions (Zhou et al., 2010; Li et al., 2011; Lynch et al., 2015). In our study, all the Env clones from DRVI01 presented mutations in the contact regions for VRC01, in particular, residues in loop D and the V5/β24 loop displayed continuous mutations (**Supplementary Table 1**).

It has been reported that single or combined mutations at position 279/281/282 were the common escape pathway of HIV-1 under immune pressure of VRC01-class antibodies (Li et al., 2011; Lynch et al., 2015). In our study, the 279E mutation was mainly observed at the first three time points, the position gradually shifted to E279K at later time points. Residues A281 was relatively constant at the first two time points, a mutation in position 281 emerged at the third time point, and the position was substituted to D/E/T in all of the isolates at three later time points. In contrast, a K282R mutation presented at the first two time points, the position K282 was relatively constant at the later four time points. Mutations and insertion within the V5/β24 loop were mainly observed in the tip of V5 loop, which had been considered to be unchangeable to neutralization sensitivity of VRC01. For Env clones from DRVI01, more substitutions and insertions in V5 loop were found at the later four time points with a trend of gradually increasing diversification over time. Almost all of Env clones presented a mutation N462T, and combined with N460 produced a PNGS, which may be a potential obstruction to the interaction between VRC01 and Env (Guo et al., 2012). Altogether, continuous mutations within loop D and the V5 loop resulted in viruses evasion from VRC01-class antibodies, more changes in loop D and the V5 loop were attended by the emergence of the A7 antibody, suggesting that the diversification in contact residues may be associated with the induction and development of A7 lineage antibodies.

In contrast, we found that almost all pseudoviruses were highly sensitive to PGT121, PGT135, 2G12, and 10E8 (most of IC_{50} < 0.5 μg/ml), which were consistent with the observed indistinguishable variants in critical residues of these bNAbs at different time points (**Table 3** and **Supplementary Table 1**).

All Env Pseudoviruses Were Resistant to DRVIA7 Lineage Antibodies

DRVIA7(A7), a VRC01-like antibody, was isolated from the patient's PBMC of 2009 time point. Longitudinal tracing of the B cell repertoires across 2006, 2008, and 2009 showed that the heavy chain of the antibody matured rapidly in 2 years, and the functional precursors of the A7 lineage heavy chain might emerge in 2006 (Kong et al., 2016). To search for the functional Env with the binding ability to the early stage A7 lineage, we tested neutralization sensitivity of Env pseudoviruses to five A7 lineage reconstituted antibodies. Our results showed that all of the Env pseudoviruses from DRVI01 were resistant to five A7 reconstituted antibodies (**Table 3**), which indicated that functional Env clones binding to the A7 lineage may be relatively rare. The results were consistent with the sequence analysis that all of the isolates from DRVI01 contained mutations in critical contact regions of VRC01 (**Supplementary Table 1**).

DISCUSSION

In a previous study, a VRC01-like antibody A7 was isolated from an HIV-1-infected individual with potent neutralization activity (Kong et al., 2016). Longitudinal analysis of B cell repertoires across 2006, 2008, and 2009 revealed that the A7 heavy chain rapidly matured within 2 years, reached peak in 2008, and declined in 2009 due to stalled light chain maturation. In this study, we isolated 155 full length Env sequences (including 68 functional Envs) from six sequential plasma samples of the patient, using SGA, to explore the characteristics of Env associated with A7 development.

TABLE 2 | Neutralization sensitivity of the Env pseudoviruses against autologous plasmas

Pseudoviruses	Autologous neutralization ID50				
	2005-7-12	2005-10-18	2006-4-6	2008-3-20	2009-12-25
2005-7-12					
0507-12	<20	172	187	413	708
0507-13	<20	138	391	206	1086
0507-14	<20	158	262	310	852
0507-15	<20	98	664	553	1407
0507-16	<20	155	318	698	759
Potency (GMTs)	<20	135	384	396	997
Breadth (n = 5)	0 (0)	5 (100%)	5 (100%)	5 (100%)	5 (100%)
2006-4-6					
0604-21	<20	<20	<20	232	603
0604-22	<20	<20	<20	258	1541
0604-23	<20	<20	<20	328	1165
0604-24	<20	<20	<20	89	1254
0604-25	<20	<20	<20	99	743
0604-26	<20	<20	<20	201	865
0604-27	<20	<20	<20	176	711
0604-28	<20	<20	<20	58	943
Potency (GMTs)	<20	<20	<20	139	810
Breadth(n = 8)	0 (0)	0 (0)	0 (0)	8 (100%)	8 (100%)
2008-3-20					
0803-11	<20	<20	<20	<20	654
0803-12	<20	<20	<20	<20	764
0803-13	<20	<20	<20	<20	773
0803-14	<20	<20	<20	<20	854
0803-15	<20	<20	<20	<20	494
0803-16	<20	<20	<20	<20	1125
Potency (GMTs)	<20	<20	<20	<20	741
Breadth(n = 6)	0 (0)	0 (0)	0 (0)	0 (0)	6 (100%)
2009-12-25					
0912-1	<20	<20	<20	<20	<20
0912-2	<20	<20	<20	<20	<20
0912-3	<20	<20	<20	<20	<20
0912-4	<20	<20	<20	<20	<20
0912-5	<20	<20	<20	<20	<20
0912-6	<20	<20	<20	<20	<20
0912-7	<20	<20	<20	<20	<20
Potency (GMTs)	<20	<20	<20	<20	<20
Breadth(n = 7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a] *The neutralizing potency is measured as ID50 in dilution of the plasmas samples. The plasmas dilution values >1000 are highlighted in red, 200∼1000 in yellow, and 20∼200 in green.* [b] *GMT: geometric mean titer.* [c] *The neutralizing breadth is calculated as the percentage of viruses neutralized with ID50 > 20.* [d] *Pseudoviruses were named by the time point + number.*

Diversification of Env was associated with the induction and maturation of bNAbs, which provided a wealth of antigenic stimulation for the immune system, and increased the probability of activating B cell precursors of bNAbs (Wibmer et al., 2013; Doria-Rose et al., 2014; Bhiman et al., 2015). In the present study, although neutralization breadth had already reached a plateau in 2005 (Kong et al., 2016), the continuous evolution of NAbs was observed to be attended by an active mutation of the virus through six time points. Gene distance and phylogenetic analysis of gp160 sequences exhibited a trend of continuous evolution. Diversification of Env gradually increased from time point 2005-10, and diversification of Env in time points 2008-3, 2009-5, and 2009-12 was significantly higher than that of the first three time points, which was consistent with emergence of A7 lineage precursors in 2006. The results indicated a continuous active interaction between the virus and the immune system in the patient.

Sequence analysis showed a shorter V1 region, lower PNGS and a lower ratio of NXT:NXS in the first five time points compared with those in time point 2009-12 and the Chinese B_database. V1V2 loop locates in the apex of the functional Env spike, and displays high amino acid variability (Wyatt et al., 1995). HIV-1 may escape from neutralization by means of a conformational mask, glycan shield, and so on (Van Gils et al., 2011). Deletion of V1V2 loop or diminishing the glycan in V1V2 could increase the neutralization sensitivity of autologous plasma and NAbs, indicating that the V1V2 loop plays an important role in the shield of the vulnerability site of the Env spike (Cao et al., 1997; Pinter et al., 2004; Bontjer et al., 2009). The Virus could escape the neutralization of NAbs by increasing the length of the V1V2 loop and number of the glycan in V1V2 (Van Gils et al., 2011; Wu et al., 2012). Some studies observed a correlation between shorter V1, lower PNGS and the induction of bNAbs, which could be explained by reducing the shield in inner vulnerability sites of the Env spike (Rademeyer et al., 2007; Bunnik et al., 2010; van den Kerkhof et al., 2013). Env V1 mutations were found to be adjacent to contact residues for CD4 and VRC01, insertions in V1 would inhibit the access to the CD4bs in the trimer (Liao et al., 2013). A shorter V1 region and lower PNGS retaining at the first five time points from DRVI01 may favor the development of Nabs.

The development of bNAbs was shown to correlate with continuous mutations directly in or adjacent to the NAbs/Env contact region, which allowed sufficient somatic hypermutation of BCR and focuses the B-cell response to the conserved vulnerability sites on Env (Sather et al., 2009; Klein et al., 2012; Liao et al., 2013). Loop D, CD4 binding loop and V5 loop play important roles in VRC01 binding with gp120 (Zhou et al., 2010; Li et al., 2011; Lynch et al., 2015). In the study, longitudinally tracking the evolution of DRVI01 *env* genes showed that more mutations presented in loop D and V5 regions over the five time points. All 155 Env clones contained three PNGS in positions 276, 460, and 463, which could obstruct the binding of the germlines of VRC01-class antibodies with the Env spike (Li et al., 2011; Guo et al., 2012; Wang et al., 2015). Previous longitudinal tracing of the A7 lineage development inferred the birth date of the A7 precursor B cells shortly before time point 2006 (Kong et al., 2016), it could be postulated that the Env variants prior to that time point may activate the precursor B cells of A7 lineage, though it could not be isolated by SGA in our study. Additionally, we observed that the diversification and mutations in loop D and the V5 region began to arise from the 2006 time point, variants with different

TABLE 3 | Neutralization sensitivity of the Env pseudoviruses against bNAbs and reconstituted A7 lineage antibodies

Pseudoviruses	IC_{50} (μg/ml)										
	PGT121	PGT135	2G12	VRC01	10E8	12A21	gDRVI01-H68+ VRC01 L	gDRVI01-H69+ VRC01 L	DRVIA7H+ gDRVIA7-L40	DRVIA7H+ gDRVIA7-L42	DRVIA7
0507-12	0.54	1.56	0.18	>50	0.72	>50	>50	>50	>50	>50	>50
0507-13	0.42	0.34	0.30	>50	0.65	>50	>50	>50	>50	>50	>50
0507-14	0.02	2.02	0.17	>50	0.71	>50	>50	>50	>50	>50	>50
0507-15	0.61	0.97	0.20	>50	0.13	>50	>50	>50	>50	>50	>50
0507-16	0.44	0.21	0.12	0.84	0.09	1.43	>50	>50	>50	>50	>50
0604-21	0.56	1.81	0.20	>50	0.15	>50	>50	>50	>50	>50	>50
0604-22	0.51	2.42	0.25	>50	0.65	>50	>50	>50	>50	>50	>50
0604-23	0.27	0.12	0.18	>50	2.29	>50	>50	>50	>50	>50	>50
0604-24	0.70	1.19	0.14	>50	3.67	>50	>50	>50	>50	>50	>50
0604-25	0.26	0.33	0.14	>50	0.32	>50	>50	>50	>50	>50	>50
0604-26	0.17	0.16	0.13	>50	0.18	>50	>50	>50	>50	>50	>50
0604-27	4.12	0.12	0.12	>50	0.81	>50	>50	>50	>50	>50	>50
0604-28	0.02	0.30	0.41	>50	0.12	>50	>50	>50	>50	>50	>50
0803-11	0.12	0.38	0.55	>50	1.10	>50	>50	>50	>50	>50	>50
0803-12	0.58	0.24	0.97	>50	0.11	>50	>50	>50	>50	>50	>50
0803-13	0.13	0.17	0.34	>50	0.59	>50	>50	>50	>50	>50	>50
0803-14	0.17	0.20	0.32	>50	0.34	>50	>50	>50	>50	>50	>50
0803-15	0.63	0.35	0.48	>50	0.61	>50	>50	>50	>50	>50	>50
0803-16	0.92	0.37	0.18	>50	0.09	>50	>50	>50	>50	>50	>50
0912-1	2.44	2.94	0.65	>50	0.30	>50	>50	>50	>50	>50	>50
0912-2	1.65	0.37	0.92	>50	0.07	>50	>50	>50	>50	>50	>50
0912-3	4.04	0.22	2.44	>50	0.19	>50	>50	>50	>50	>50	>50
0912-4	1.37	3.77	0.28	>50	0.03	>50	>50	>50	>50	>50	>50
0912-5	0.18	0.21	0.17	>50	0.19	>50	>50	>50	>50	>50	>50
0912-6	0.58	1.37	0.40	>50	0.31	>50	>50	>50	>50	>50	>50
0912-7	1.49	1.90	0.41	>50	0.18	>50	>50	>50	>50	>50	>50

[a] *The neutralizing potency is measured as IC_{50} in μg/ml of the monoclonal antibodies. Values < 0.2μg/ml are highlighted in red, 0.2∼2μg/ml in yellow, and 2∼20μg/ml in green.* [b] *The characteristics and neutralization data of A7 lineage reconstituted Abs refer to the study from Kong et al. (2016).* [c] *Pseudoviruses were named by the time point + number.*

modes alternatively appeared over time. The results indicated that viruses escaping neutralization of VRC01-class antibodies, by means of mutating NAbs contact residues, could drive the breadth of NAbs.

The robust immune responses driving continuous escape mutants facilitated the development of bNAbs (Liao et al., 2013; Wibmer et al., 2013; Gao et al., 2014; Bonsignori et al., 2016). In the present study, neutralization sensitivity of the Env isolates against autologous plasma and a few well-known bNabs was analyzed to observe virus-antibody interactions. The plasma of DVRI01 presented potent and broad neutralization activity (Hu et al., 2012; Kong et al., 2016). As expected, almost all Env isolates from the patient escaped neutralization by concurrent autologous plasma. The results indicated that the strong autologous neutralizing selection pressure continuously drove the viruses to escape. Neutralization sensitivity of the Env pseudoviruses against a few bNAbs showed that all Env clones were VRC01-resistant, suggesting the presence of strong immune pressure from the VRC01-class antibodies in the patient. In contrast, PGT121, PGT135, 10E8, and 2G12 which target the glycans in the V3 region, MPER, and the glycans in outer gp120, respectively, could potently neutralize all Env clones, which were in line with the conserved critical residues of these bNAbs epitopes over the study period, indicating a lack of immune pressure of the above four bNAbs in the patient.

The functional Env clones binding germline precursors of bNAbs have been considered as potential virus strains initiating the development of bNAbs. In the present study, however, all Env pseudoviruses derived from DRVI01 plasma were proven to be neutralization-resistant toward the five reconstituted A7 lineage antibodies, suggesting that the virus strains associated with the development of the A7 lineage were not dominant. Our previous analysis of the germline gene usage also displayed that IgHV1-2, the germline family of DRVIA7 heavy chain, were significantly lower than IgHV4-34 and IgHV4-39 across the 2006, 2008, and 2009 time points, indicating that DRVIA7 did not constitute a major lineage within the repertoire (Kong et al., 2016). Dynamic antibody evolution revealed that A7 lineage precursors emerged in 2006, but all viruses isolated before 2006 were resistant to A7 lineage antibodies, suggesting that immune pressure of VRC01-class antibodies already presented prior to 2006. The results were consistent with an Env sequences analysis which

showed many mutations already presented in loop D and V5 at the two 2005 time points. Therefore, the immune pressure of VRC01-class antibodies presented prior to the timeframe studied could result in failure in isolating the Env clones capable of binding A7 lineages.

In summary, we acquired 155 intact *env* sequences from a Chinese chronically HIV-1-infected individual with potent neutralization activity. Both the sequence and neutralization analysis showed that the gradually increasing diversification of the Env sequences was associated with the development of the A7 lineage; the robust neutralization activity of plasmas and the escaped mutants from autologous plasmas were consistent with more mutations in the contact region of Nabs, which suggests a continuous co-evolution of Env and Nabs. Additionally, sequences analysis observed a few characteristics that could facilitate the recognition of CD4bs antibodies, which contained shorter V1, lower PNGS and a ratio of NXT:NXS at the first five time points.

There were several limitations in the study. First, the subject DRVI01 was infected over 10 years, the early samples of infection were unavailable for the study. Second, the precursor of the heavy chain of the A7 lineage was inferred to emerge before 2006, but the unmutated common ancestor (UCA) had not been identified in previous studies. Third, we could not identify the Env variants capable of binding the A7 lineage antibodies.

ETHICS STATEMENT

The study was reviewed and approved by the Institutional Review Board of the National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention. The subject provided written informed consent before blood and data collection.

AUTHOR CONTRIBUTIONS

DZ, KH, and YS designed the study, analyzed the data, and edited the manuscript. DZ, SZ, YH, and XH performed the *env* cloning, sequencing, and phylogenetic analysis. DZ, JH, SZ, XTH, and LR prepared the pseudoviruses and performed the neutralization assays. YS, LM, and KH developed the cohort and collected the samples.

FUNDING

This work was supported by a grant from the National Natural Science Foundation of China (81172809 and 31411130194), the National Major Project for Infectious Disease Control and Prevention (2016ZX10001-008 and 2018ZX10731-101), the SKLID key project (2016SKLID602), and the International Cooperation Grant from the Ministry of Science and Technology of China (2016YFE0107600).

ACKNOWLEDGMENTS

We are grateful to the Vaccine Research Center of NIAID at NIH for providing the RSC3/ΔRSC3 probes and VRC01. We thank the CAVD HIV Specimen Cryorepository (HSC) for their contribution of HIV-1 Env-pseudotyped viruses for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01096/full#supplementary-material

REFERENCES

Barre-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., Gruest, J., et al. (1983). Isolation of a t-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (aids). *Science* 220, 868–871. doi: 10.1126/science.6189183

Bhiman, J. N., Anthony, C., Doriarose, N. A., Karimanzira, O., Schramm, C. A., Khoza, T., et al. (2015). Viral variants that initiate and drive maturation of v1v2-directed HIV-1 broadly neutralizing antibodies. *Nat. Med.* 21, 1332–1336. doi: 10.1038/nm.3963

Bonsignori, M., Liao, H. X., Gao, F., Williams, W. B., Alam, S. M., Montefiori, D. C., et al. (2017). Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development. *Immunol. Rev.* 275, 145–160. doi: 10.1111/imr.12509

Bonsignori, M., Zhou, T., Sheng, Z., Chen, L., Gao, F., Joyce, M. G., et al. (2016). Maturation pathway from germline to broad HIV-1 neutralizer of a CD4-mimic antibody. *Cell* 165, 449–463. doi: 10.1016/j.cell.2016.02.022

Bontjer, I., Land, A., Eggink, D., Verkade, E., Tuin, K., Baldwin, C., et al. (2009). Optimization of human immunodeficiency virus type 1 envelope glycoproteins with v1/v2 deleted, using virus evolution. *J. Virol.* 83, 368–383. doi: 10.1128/JVI.01404-08

Buchacher, A., Predl, R., Strutzenberger, K., Steinfellner, W., and Jungbauer, A. (1994). Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res. Hum. Retroviruses* 10, 359–369. doi: 10.1089/aid.1994.10.359

Bunnik, E. M., Euler, Z., Welkers, M. R., Boesernunnink, B. D., Grijsen, M. L., Prins, J. M., et al. (2010). Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. *Nat. Med.* 16, 995–997. doi: 10.1038/nm.2203

Bunnik, E. M., Pisas, L., Van Nuenen, A. C., and Schuitemaker, H. (2008). Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype b human immunodeficiency virus type 1 infection. *J. Virol.* 82, 7932–7941. doi: 10.1128/JVI.00757-08

Burton, D. R. (2019). Advancing an HIV vaccine; advancing vaccinology. *Nat. Rev. Immunol.* 19, 77–78. doi: 10.1038/s41577-018-0103-6

Burton, D. R., and Hangartner, L. (2016). Broadly neutralizing antibodies to HIV and their role in vaccine design. *Annu. Rev. Immunol.* 34, 635–659. doi: 10.1146/annurev-immunol-041015-055515

Cao, J., Sullivan, N., Desjardin, E., Parolin, C., Robinson, J., Wyatt, R., et al. (1997). Replication and neutralization of human immunodeficiency virus type 1 lacking the v1 and v2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* 71, 9808–9812.

Conley, A. J., Kessler, J. A., Boots, L. J., Tung, J. S., Arnold, B. A., Keller, P. M., et al. (1994). Neutralization of divergent human immunodeficiency virus type 1 variants and primary isolates by IAM-41-2f5, an anti-gp41 human monoclonal antibody. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3348–3352. doi: 10.1073/pnas.91.8.3348

Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984). The CD4 (t4) antigen is an essential component of the receptor for the aids retrovirus. *Nature* 312, 763–767. doi: 10.1038/312763a0

Doria-Rose, N. A., Schramm, C. A., Gorman, J., Moore, P. L., Bhiman, J. N., DeKosky, B. J., et al. (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* 509, 55–62. doi: 10.1038/nature13036

Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., et al. (1983). Isolation of human t-cell leukemia virus in acquired immune deficiency syndrome (aids). *Science* 220, 865–867. doi: 10.1126/science.6601823

Gao, F., Bonsignori, M., Liao, H. X., Kumar, A., Xia, S. M., Lu, X., et al. (2014). Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell* 158, 481–491. doi: 10.1016/j.cell.2014.06.022

Gavel, Y., and Heijne, G. V. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng. Des. Sel.* 3, 433–442. doi: 10.1093/protein/3.5.433

Gottlieb, M. S., and Schroff, R. (1981). *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N. Engl. J. Med.* 305, 1425–1431. doi: 10.1056/NEJM198112103052401

Guo, D., Shi, X., Arledge, K. C., Song, D., Jiang, L., Fu, L., et al. (2012). A single residue within the V5 region of HIV-1 envelope facilitates viral escape from the broadly neutralizing monoclonal antibody VRC01. *J. Biol. Chem.* 287, 43170–43179. doi: 10.1074/jbc.M112.399402

Haynes, B. F., and Mascola, J. R. (2017). The quest for an antibody-based HIV vaccine. *Immunol. Rev.* 275, 5–10. doi: 10.1111/imr.12517

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Hraber, P., Seaman, M. S., Bailer, R. T., Mascola, J. R., Montefiori, D. C., and Korber, B. T. (2014). Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* 28, 163–169. doi: 10.1097/QAD.0000000000000106

Hu, X., Hong, K., Zhao, C., Zheng, Y., Ma, L., Ruan, Y., et al. (2012). Profiles of neutralizing antibody response in chronically human immunodeficiency virus type 1 clade B′-infected former plasma donors from China naive to antiretroviral therapy. *J. Gen. Virol.* 93(Pt_10), 2267–2278. doi: 10.1099/vir.0.043802-0

Huang, J., Doria-Rose, N. A., Longo, N. S., Laub, L., Lin, C. L., Turk, E., et al. (2013). Isolation of human monoclonal antibodies from peripheral blood B cells. *Nat. Protoc.* 8, 1907–1915. doi: 10.1038/nprot.2013.117

Huang, J., Kang, B. H., Ishida, E., Zhou, T., Griesman, T., Sheng, Z., et al. (2016). Identification of a CD4-binding-site antibody to HIV that evolved near-pan neutralization breadth. *Immunity* 45, 1108–1121. doi: 10.1016/j.immuni.2016.10.027

Huang, J., Ofek, G., Laub, L., Louder, M. K., Doria-Rose, N. A., Longo, N. S., et al. (2012). Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* 491, 406–412. doi: 10.1038/nature11544

Julg, B., Liu, P. T., Wagh, K., Fischer, W. M., Abbink, P., Mercado, N. B., et al. (2017). Protection against a mixed SHIV challenge by a broadly neutralizing antibody cocktail. *Sci. Transl. Med.* 9:eaao4235. doi: 10.1126/scitranslmed.aao4235

Kaplan, H. A., Welply, J. K., and Lennarz, W. J. (1987). Oligosaccharyl transferase: the central enzyme in the pathway of glycoprotein assembly. *Biochim. Biophys. Acta* 906, 161–173.

Klein, F., Gaebler, C., Mouquet, H., Sather, D. N., Lehmann, C., Scheid, J. F., et al. (2012). Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J. Exp. Med.* 209, 1469–1479. doi: 10.1084/jem.20120423

Kong, L., Ju, B., Chen, Y., He, L., Ren, L., Liu, J., et al. (2016). Key gp120 glycans pose roadblocks to the rapid development of VRC01-class antibodies in an HIV-1-infected Chinese donor. *Immunity* 44, 939–950. doi: 10.1016/j.immuni.2016.03.006

Kwong, P. D., and Mascola, J. R. (2018). HIV-1 vaccines based on antibody identification, B cell ontogeny, and epitope structure. *Immunity* 48, 855–871. doi: 10.1016/j.immuni.2018.04.029

Li, M., Gao, F., Mascola, J. R., Stamatatos, L., Polonis, V. R., Koutsoukos, M., et al. (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype b infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* 79, 10108–10125.

Li, Y., O'Dell, S., Walker, L. M., Wu, X., and Mascola, J. R. (2011). Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. *J. Virol.* 85, 8954–8967. doi: 10.1128/JVI.00754-11

Liao, H. X., Lynch, R., Zhou, T., Gao, F., Alam, S. M., Boyd, S. D., et al. (2013). Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496, 469–476. doi: 10.1038/nature12053

Lu, C. L., Murakowski, D. K., Bournazos, S., Schoofs, T., Sarkar, D., Halper-Stromberg, A., et al. (2016). Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 in vivo. *Science* 352, 1001–1004. doi: 10.1126/science.aaf1279

Lynch, R. M., Wong, P., Tran, L., O'Dell, S., Nason, M. C., Li, Y., et al. (2015). HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* 89, 4201–4213. doi: 10.1128/JVI.03608-14

Moore, P. L. (2018). The neutralizing antibody response to the HIV-1 env protein. *Curr. HIV Res.* 16, 21–28. doi: 10.2174/1570162X15666171124122044

Moore, P. L., Gray, E. S., and Morris, L. (2009a). Specificity of the autologous neutralizing antibody response. *Curr. Opin. HIV AIDS* 4, 358–363. doi: 10.1097/COH.0b013e32832ea7e8

Moore, P. L., Ranchobe, N., Lambson, B. E., Gray, E. S., Cave, E., Abrahams, M. R., et al. (2009b). Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog.* 5:e1000598. doi: 10.1371/journal.ppat.1000598

Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., et al. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* 67, 6642–6647.

Pinter, A., Honnen, W. J., He, Y., Gorny, M. K., Zolla-Pazner, S., and Kayman, S. C. (2004). The V1/V2 domain of gp120 is a global regulator of the sensitivity of primary human immunodeficiency virus type 1 isolates to neutralization by antibodies commonly induced upon infection. *J. Virol.* 78, 5205–5215.

Rademeyer, C., Moore, P. L., Taylor, N., Martin, D. P., Choge, I. A., Gray, E. S., et al. (2007). Genetic characteristics of HIV-1 subtype c envelopes inducing cross-neutralizing antibodies. *Virology* 368, 172–181.

Richman, D. D., Wrin, T., Little, S. J., and Petropoulos, C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4144–4149. doi: 10.1073/pnas.0630530100

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–57. doi: 10.1126/science.288.5463.55d

Rusert, P., Krarup, A., Magnus, C., Brandenberg, O. F., Weber, J., Ehlert, A. K., et al. (2011). Interaction of the gp120 v1v2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies. *J. Exp. Med.* 208, 1419–1433. doi: 10.1084/jem.20110196

Sather, D. N., Armann, J., Ching, L. K., Mavrantoni, A., Sellhorn, G., Caldwell, Z., et al. (2009). Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *J. Virol.* 83, 757–769. doi: 10.1128/JVI.02036-08

Simek, M. D., Rida, W., Priddy, F. H., Pung, P., Carrow, E., Laufer, D. S., et al. (2009). Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J. Virol.* 83, 7337–7348. doi: 10.1128/JVI.00110-09

Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., and Sullivan, N. (1996). Human monoclonal antibody 2g12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* 70, 1100–1108.

van den Kerkhof, T. L., Feenstra, K. A., Euler, Z., van Gils, M. J., Rijsdijk, L. W., Boeser-Nunnink, B. D., et al. (2013). HIV-1 envelope glycoprotein signatures that correlate with the development of cross-reactive neutralizing activity. *Retrovirology* 10:102. doi: 10.1186/1742-4690-10-102

Van Gils, M. J., Bunnik, E. M., Boeser-Nunnink, B. D., Burger, J. A., Terlouw-Klein, M., Verwer, N., et al. (2011). Longer v1v2 region with increased number of potential n-linked glycosylation sites in the HIV-1 envelope glycoprotein protects against HIV-specific neutralizing antibodies. *J. Virol.* 85, 6986–6995. doi: 10.1128/JVI.00268-11

Van Gils, M. J., and Sanders, R. W. (2013). Broadly neutralizing antibodies against HIV-1: templates for a vaccine. *Virology* 435, 46–56. doi: 10.1016/j.virol.2012.10.004

Walker, L. M., Huber, M., Doores, K. J., Falkowska, E., Pejchal, R., Julien, J. P., et al. (2011). Broad neutralization coverage of HIV by multiple

highly potent antibodies. *Nature* 477, 466–470. doi: 10.1038/nature 10373

Walker, L. M., Phogat, S. K., Chan-Hui, P. Y., Wagner, D., Phung, P., Goss, J. L., et al. (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326, 285–289. doi: 10.1126/science. 1178746

Wang, W., Zirkle, B., Nie, J., Ma, J., Gao, K., Chen, X. S., et al. (2015). N463 glycosylation site on v5 loop of a mutant gp120 regulates the sensitivity of HIV-1 to neutralizing monoclonal antibodies VRC01/03. *J. Acquir. Immune Defic. Syndr.* 69, 270–277. doi: 10.1097/QAI.0000000000000595

Wibmer, C. K., Bhiman, J. N., Gray, E. S., Tumba, N., Abdool Karim, S. S., Williamson, C., et al. (2013). Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. *PLoS Pathog.* 9:e1003738. doi: 10.1371/journal.ppat.1003738

Wu, X., and Kong, X. P. (2016). Antigenic landscape of the HIV-1 envelope and new immunological concepts defined by HIV-1 broadly neutralizing antibodies. *Curr. Opin. Immunol.* 42, 56–64. doi: 10.1016/j.coi.2016.05.013

Wu, X., Wang, C., O'Dell, S., Li, Y., and Mascola, J. R. (2012). Selection pressure on HIV-1 envelope by broadly neutralizing antibodies to the conserved CD4-binding site. *J. Virol.* 86, 5844–5856. doi: 10.1128/JVI. 07139-11

Wyatt, R., Moore, J., Accola, M., Desjardin, E., Robinson, J., and Sodroski, J. (1995). Involvement of the v1/v2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* 69, 5723–5733. doi: 10.1016/0166-0934(95) 00053-W

Zhou, T., Georgiev, I., Wu, X., Yang, Z. Y., Dai, K., Finzi, A., et al. (2010). Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* 329, 811–817. doi: 10.1126/science.1192819

Zhou, T., Lynch, R. M., Chen, L., Acharya, P., Wu, X., Doria-Rose, N. A., et al. (2015). Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell* 161, 1280–1292. doi: 10.1016/j.cell.2015. 05.007

Zhu, J., Ofek, G., Yang, Y., Zhang, B., and Kwong, P. D. (2013). Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6470–6475. doi: 10.1073/pnas.1219320110

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 June 2019
doi: 10.3389/fmicb.2019.01340

Phylogeographic Analyses Reveal the Early Expansion and Frequent Bidirectional Cross-Border Transmissions of Non-pandemic HIV-1 Subtype B Strains in Hispaniola

*Gonzalo Bello[1], Ighor Arantes[1], Vincent Lacoste[2], Marlene Ouka[3], Jacques Boncy[4], Raymond Césaire[3], Bernard Liautaud[5], Mathieu Nacher[6] and Georges Dos Santos[3]**

[1]Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, [2]Laboratoire des Interactions Virus-Hôtes, Institut Pasteur de la Guyane, Cayenne, French Guiana, [3]Virology Laboratory, EA 4537, Martinique University Hospital, Fort de France, Martinique, [4]Laboratoire National de Santé Publique, Ministère de la Santé Publique et de la Population, Port-au-Prince, Haiti, [5]Training Unit, GHESKIO Centers, Port-au-Prince, Haiti, [6]Coordination Régionale de la lutte contre le VIH (COREVIH) and Centre d'Investigation Clinique—CIC INSERM 1424, Centre Hospitalier de Cayenne "Andrée Rosemon", Cayenne, French Guiana

The human immunodeficiency virus-type 1 (HIV-1) subtype B has probably been circulating on the island of Hispaniola since the 1960s, but information about the early viral history on this Caribbean island is scarce. In this study, we reconstruct the dissemination dynamics of early divergent non-pandemic subtype B lineages (designated B_{CAR}) on Hispaniola by analyzing a country-balanced dataset of HIV-1 B_{CAR} *pol* sequences from Haiti (n = 103) and the Dominican Republic (n = 123). Phylogenetic analyses supported that B_{CAR} strains from Haiti and the Dominican Republic were highly intermixed between each other, although the null hypothesis of completely random mixing was rejected. Bayesian phylogeographic analyses placed the ancestral B_{CAR} virus in Haiti and the Dominican Republic with the same posterior probability support. These analyses estimate frequent viral transmissions between Haiti and the Dominican Republic since the early 1970s onwards, and the presence of local B_{CAR} transmission networks in both countries before first AIDS cases was officially recognized. Demographic reconstructions point that the B_{CAR} epidemic in Hispaniola grew exponentially until the 1990s. These findings support that the HIV-1 epidemics in Haiti and the Dominican Republic have been connected by a recurrent bidirectional viral flux since the initial phase, which poses a great challenge in tracing the geographic origin of the B_{CAR} epidemic within Hispaniola using only genetic data. These data also reinforce the notion that prevention programs have successfully reduced the rate of new HIV-1 transmissions in Hispaniola since the end of the 1990s.

Keywords: HIV-1, subtype B, non-pandemic, origin, phylodynamics, Haiti, Dominican Republic

OPEN ACCESS

Edited by:
Joris Hemelaar, University of Oxford, United Kingdom

Reviewed by:
Shigeyoshi Harada, National Institute of Infectious Diseases (NIID), Japan
Maja Stanojevic, University of Belgrade, Serbia
Bram Vrancken, KU Leuven, Belgium

****Correspondence:***
Georges Dos Santos
georges.dos-santos@chu-martinique.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *27 September 2018*
Accepted: *29 May 2019*
Published: *26 June 2019*

Citation:
Bello G, Arantes I, Lacoste V, Ouka M, Boncy J, Césaire R, Liautaud B, Nacher M and Dos Santos G (2019) Phylogeographic Analyses Reveal the Early Expansion and Frequent Bidirectional Cross-Border Transmissions of Non-pandemic HIV-1 Subtype B Strains in Hispaniola. Front. Microbiol. 10:1340. doi: 10.3389/fmicb.2019.01340

INTRODUCTION

The island of Hispaniola, shared by Haiti and the Dominican Republic, included in 2016 around 217,000 people living with the human immunodeficiency virus-type 1 (HIV-1), the etiologic agent of the acquired immunodeficiency syndrome (AIDS) (UNAIDS, 2013, 2017). Initially, AIDS cases recognized in Haiti (Pape et al., 1983) and the Dominican Republic (Koenig et al., 1987) mostly involved men who had sex with men, but these countries now have generalized epidemics predominantly driven by heterosexual sex (Figueroa, 2008). Despite the high HIV prevalence rates reported for the general population in Haiti (2.1%: 1.9–2.3%) and the Dominican Republic (1.0%: 0.7–1.4%) in 2016, significant declines of the HIV incidence rates were observed in both countries over the last decade (UNAIDS, 2013, 2017).

The HIV-1 epidemic in Hispaniola is dominated by subtype B (Nadai et al., 2009; Myers et al., 2012; Lopez et al., 2015). Genetic evidence suggests that the subtype B epidemic in Hispaniola, as in many other Caribbean countries, is mainly driven by the transmission of multiple early divergent non-pandemic subtype B lineages (designated "B_{CAR}"), although the worldwide disseminated "$B_{PANDEMIC}$" lineage also circulates (Cabello et al., 2014, 2015; Divino et al., 2016; Bello et al., 2018). The early dissemination dynamics of the HIV-1 subtype B epidemics between Haiti and the Dominican Republic remain largely unknown. Previous evolutionary analyses pointed that subtype B probably entered in the Americas through Haiti around the mid-1960s and then moved to other countries (Gilbert et al., 2007; Junqueira et al., 2011; Worobey et al., 2016), but no or very few ($n < 15$) subtype B sequences from the Dominican Republic were included in those studies. Another study with a large number of HIV-1 sequences from the Dominican Republic has shown that subtype B virus circulates in this country since the early 1960s (Lopez et al., 2015), a result comparable to the estimated age of the Haitian epidemic. This study, however, may have traced the age of the Haitian subtype B ancestor rather than of the Dominican one, given that it assumed that all subtype B infections in the Dominican Republic resulted from a single introduction and that the hypothesis of multiple independent viral introductions from Haiti was not formally tested.

The objective of this work was to reconstruct the early spatiotemporal dynamics of dissemination and demographic history of non-pandemic HIV-1 B_{CAR} lineages on the island of Hispaniola and to estimate the current degree of geographic compartmentalization of the B_{CAR} epidemic between Haiti and the Dominican Republic.

MATERIALS AND METHODS

HIV-1 Subtype B *pol* Haitian Sequences

A total of 127 new HIV-1 subtype B *pol* sequences covering the complete protease (PR) and the first part of the reverse transcriptase (RT) regions (nucleotides 2253 to 3275 of reference strain HXB2) were obtained from adult patients at Port-au-Prince, Haiti, who underwent HIV genotyping tests at the Virology Laboratory of the University Hospital of Martinique (Fort-de-France, Martinique) between 2009 and 2014. All patients provided written informed consent and samples were anonymized as recommended in the study protocol approved by the Comité National d'Ethique du Ministère de la Santé Publique et de la Population de Haiti (13/07/2009). Only one sequence per subject was selected and all sites associated with major antiretroviral drug resistance were removed.

HIV-1 Subtyping and Lineage Assignment

The subtype initially assigned to new Haitian sequences by the REGA HIV subtyping tool v.2 (de Oliveira et al., 2005) was confirmed by performing a maximum likelihood (ML) phylogenetic analysis with HIV-1 group M subtype reference sequences (**Supplementary Table S1**). We also tested for recombination using the Recombination Detection Program (RDP) v4.9 (Martin et al., 2005) with the default settings. Only statistically significant ($p < 0.05$) events supported by at least two methods were considered. HIV-1 subtype B *pol* sequences from Haiti were then aligned with subtype B sequences representative of the $B_{PANDEMIC}$ and the B_{CAR} clades (**Supplementary Table S2**) selected from a previous study (Cabello et al., 2014) and subjected to ML phylogenetic analysis for lineage classification. The ML trees were reconstructed with the PhyML program (Guindon et al., 2010) using an online web server (Guindon et al., 2005) under the best fit nucleotide substitution model selected with the SMS tool (Lefort et al., 2017), the SPR branch-swapping algorithm of heuristic tree search, and the approximate likelihood-ratio test (aLRT) (Anisimova and Gascuel, 2006) of reliability tree topology. The ML trees were visualized using the FigTree v1.4 program (Rambaut, 2009).

Analysis of Population Subdivision by Country

The HIV-1 B_{CAR} *pol* sequences from Haiti identified here were aligned with B_{CAR} sequences from Haiti ($n = 12$) and the Dominican Republic ($n = 123$) previously characterized (Cabello et al., 2014). A Bayesian phylogenetic tree of the B_{CAR} dataset from Hispaniola was reconstructed under the best-fit nucleotide substitution model (GTR + I + G) using the MrBayes program (Ronquist et al., 2012). Two chains were run for 25×10^6 generations, and stationarity (constant mean and variance of trace plots) and good mixing (effective sample size > 200) for all parameter estimates were assessed using TRACER v1.7 (Rambaut et al., 2018). The degree of phylogenetic mixing of B_{CAR} sequences obtained from both countries was then quantified using the BaTS program (Parker et al., 2008), which estimates phylogeny-trait associations using the Association Index (AI) (Wang et al., 2001), the Parsimony Score (PS) (Wang et al., 2001), and the maximum clade (MC) statistics. Results were considered significant for $p < 0.01$.

Phylodynamic Analyses

The evolutionary rate, the age of the most recent common ancestor (T_{MRCA}), the spatial diffusion pattern, and the rate of population growth (r, year^{-1}) of B_{CAR} lineages in Hispaniola

were jointly estimated using a Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.8 (Drummond et al., 2002; Drummond and Rambaut, 2007). Regression analyses using the TempEst program (Rambaut et al., 2016) revealed that the B_{CAR} *pol* dataset compiled here did not contain a sufficient temporal signal for reliable time-scale estimations [*X*-intercept (T_{MRCA}) < 1910]. Thus, Bayesian MCMC analyses were performed using a relaxed uncorrelated lognormal molecular clock model (Drummond et al., 2006) with a uniform prior distribution on the substitution rate ($1.7–3.0 \times 10^{-3}$ subst./site/year), based on previous estimates (Hue et al., 2005; Zehender et al., 2010; Chen et al., 2011; Mendoza et al., 2014). Migration events were reconstructed using a reversible discrete phylogeographic model (Lemey et al., 2009) with a CTMC rate reference prior (Ferreira and Suchard, 2008). The number of location transitions (viral migrations between countries) throughout the evolutionary history was estimated using Markov jump counts (O'Brien et al., 2009). Changes in effective population size through time (*Ne*) were estimated using the non-parametric Bayesian skyline (BSKL) (Drummond et al., 2005) and Bayesian Skygrid (BSKG) (Gill et al., 2013) models. Estimates of the *r* were obtained under the best-fit parametric model selected using the log marginal likelihood estimation (MLE) based on the generalized stepping-stone sampling (GSS) method (Baele et al., 2016). Six MCMC chains were run for 200 million generations and then combined to ensure stationarity and good mixing as described above. The MC credibility (MCC) tree was summarized with TreeAnnotator v1.8 and visualized using the FigTree v1.4 program.

Statistical Analyses

Gender and age group of Haitian individuals infected with different subtype B lineages were compared using Fisher's exact test or χ^2 implemented in Stata 13 software. Statistical significance was defined as $p < 0.05$.

RESULTS

All new HIV-1 *pol* sequences from Haiti obtained here ($n = 127$) were confirmed as non-recombinant subtype B by ML phylogenetic analysis (**Supplementary Figure S1**) and RDP recombination analysis (data not shown). The HIV-1 subtype B Haitian sequences were combined with viral strains representative of the B_{CAR} diversity in different Caribbean islands ($n = 200$) and of the $B_{PANDEMIC}$ diversity in the US and France ($n = 300$) (**Supplementary Table S2**) as previously characterized (Cabello et al., 2014). The ML phylogenetic analysis revealed that most subtype B Haitian sequences ($n = 91$, 72%) were intermixed among basal non-pandemic B_{CAR} lineages, whereas the remaining ones ($n = 36$, 28%) branched within the well-supported (aLRT = 0.89) $B_{PANDEMIC}$ clade (**Figure 1**). Analysis of the epidemiological characteristics of Haitian subjects showed that both B_{CAR} and $B_{PANDEMIC}$ viral lineages circulated among males and females of different age groups (**Supplementary Table S3**). No significant differences were observed in the frequency of subtype B lineages according to gender ($p = 0.82$) or age group ($p = 0.45$), although subjects infected with B_{CAR} strains had a relative younger mean age (36.0 years) as compared with those infected with the $B_{PANDEMIC}$ clade (40.1 years) ($p = 0.03$).

A closer inspection of the phylogenetic relationship among B_{CAR} sequences confirm that sequences from Haiti and the Dominican Republic were highly intermixed with each other and are usually basal to sequences from other Caribbean islands (**Supplementary Figure S2**). Most sequences from other well-sampled islands like Jamaica and Trinidad and Tobago, by contrast, branched in country-specific subclades that were nested within the Hispaniola B_{CAR} diversity. Very few sequences from Hispaniola (<1%) branched within Trinidadian and Jamaican clusters, confirming that most B_{CAR} infections in Hispaniola resulted from internal viral dissemination and not from re-introductions of viral strains from other Caribbean islands.

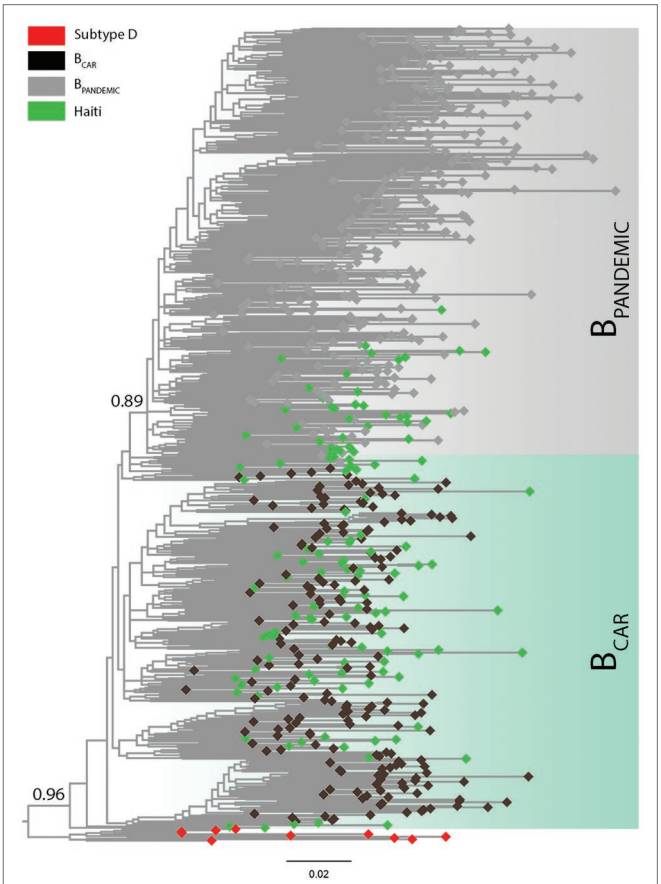


FIGURE 1 | Classification of the HIV-1 subtype B sequences from Haiti among pandemic ($B_{PANDEMIC}$) and non-pandemic (B_{CAR}) lineages. ML phylogenetic tree of HIV-1 subtype B *pol* sequences (~1,000 nt) from Haiti ($n = 127$; green tips) together with representative sequences of the $B_{PANDEMIC}$ (US = 165, France = 135; gray tips) and the B_{CAR} (Caribbean = 200; black tips) lineages. Node support (aLRT) for subtype B and $B_{PANDEMIC}$ monophyletic groups are indicated. Shaded boxes highlight the position of the B_{CAR} and $B_{PANDEMIC}$ lineages. Tree was rooted using HIV-1 subtype D reference sequences (red tips). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

To investigate the phylogeographic structure of the HIV-1 epidemic within Hispaniola, B_{CAR} *pol* sequences from Haiti here identified (n = 91) were combined with Haitian (n = 12) and Dominican (n = 123) B_{CAR} *pol* sequences identified in a previous study (Cabello et al., 2014) and analyzed using BaTS. Analyses of population subdivision rejected the null hypothesis of panmixis (i.e., complete intermixing of sequences from Haiti and the Dominican Republic) (**Supplementary Table S4**), demonstrating that despite frequent viral intermixing between both countries, the geographic subdivision of the HIV-1 B_{CAR} sequences from Hispaniola was greater than expected by chance.

The same dataset of B_{CAR} *pol* sequences from Haiti (n = 103) and the Dominican Republic (n = 123) was then subjected to Bayesian phylogeographic analyses. Reconstruction of the spatiotemporal dissemination dynamic traced the T_{MRCA} for the HIV-1 B_{CAR} epidemic in Hispaniola at 1967 (95% HPD: 1961–1972), but failed to uncover its precise epicenter (**Figure 2**). After combining six independent Bayesian MCMC runs, the root location of the HIV-1 B_{CAR} ancestor in Hispaniola was traced with equal probability [posterior state probability (PSP) = 0.50] to Haiti and to the Dominican Republic. The difficulty to trace the location of the B_{CAR} root into a single country was also evidenced when the results obtained from independent Bayesian MCMC runs were visualized separately (**Supplementary Table S5**). Quantification of B_{CAR} flux between countries using Markov jump counts support a mean of 18 viral migrations from Haiti to the Dominican Republic and 9 viral migration events from the Dominican Republic to Haiti. These viral migrations started in the early 1970s and were homogenously distributed between the mid-1970s and the mid-2000s (**Supplementary Figure S3**). The Bayesian phylogeographic analysis also revealed 20 country-specific (14 Dominican and 6 Haitian) B_{CAR} monophyletic subclades with relative high node support [posterior probability (PP) > 0.70] (**Figure 2**). The Dominican B_{CAR} subclades mostly arose between the mid-1970s and the early 1980s (**Supplementary Table S6**) and together comprised 55% (n = 68) of B_{CAR} sequences from the country.

HIV-1 B_{CAR} sequences from Haiti (n = 103) and the Dominican Republic (n = 123) were finally used to reconstruct the demographic history of this viral epidemic in Hispaniola. Reconstruction of population dynamics with the BSKL coalescent-based model suggested that the B_{CAR} epidemic in Hispaniola experienced an initial phase of fast exponential growth until the beginning of the 1990s, followed by a stabilization of the *Ne* (**Figure 3A**). The BSKG model, however, supported a longer exponential growth phase until the late 1990s, followed by an epidemic decline that extended until

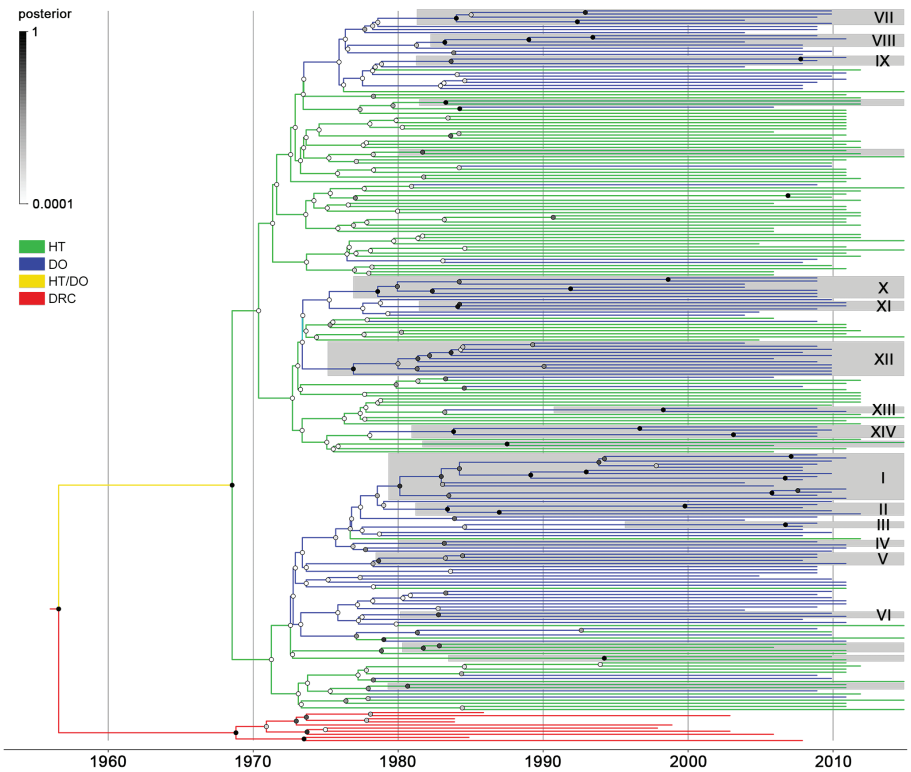


FIGURE 2 | Spatiotemporal dissemination of HIV-1 B_{CAR} lineages in Hispaniola. Time-scaled Bayesian MCC tree of HIV-1 B_{CAR} *pol* sequences from Haiti (n = 103) and the Dominican Republic (n = 123) combined with subtype D reference sequences from the Democratic Republic of Congo (DRC; n = 10). Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend on the left. Shaded boxes highlight the position of B_{CAR} clades only composed by sequences from Haiti or the Dominican Republic (identified by numbers) and that displayed both high clade (PP ≥ 0.70) and location state (PSP ≥ 0.90) node supports. Circles at internal nodes are colored according to the corresponding PP node support as indicated in the legend on the left. Branch lengths are depicted in units of time (years). The tree was rooted under the assumption of a relaxed molecular clock.

the most recent coalescent event around the mid-2000s (**Figure 3B**). The UNAIDS epidemiological data, in agreement with the BSKG model, indicate a growth in the number of new HIV infections in Hispaniola until the early 2000s and a subsequent phase of decline extending until the most recent date (**Figure 3C**). To estimate the growth rate of the B_{CAR} epidemic at the initial phase, three parametric coalescent models (logistic, exponential, and expansion) were compared. According to the best-fit logistic growth coalescent model (**Supplementary Table S7**), the mean growth rate of the B_{CAR} epidemic during the first decades of expansion in Hispaniola was 0.50 (95% HPD: 0.37–0.65).

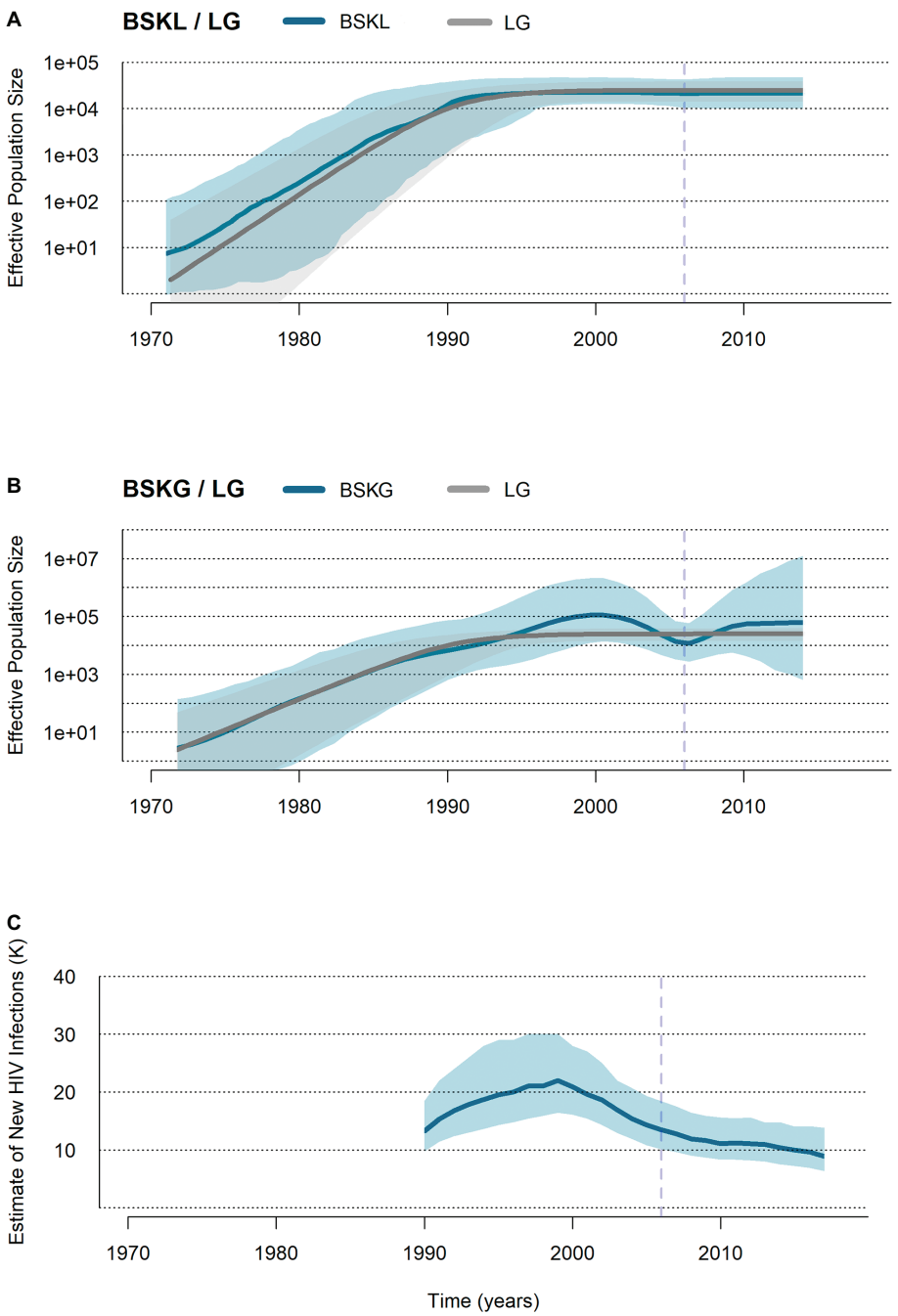


FIGURE 3 | Population and epidemiological dynamics of the HIV-1 B_{CAR} epidemic in Hispaniola. **(A,B)** Plots showing the median (solid blue lines) and the 95% HPD intervals (dashed blue areas) estimates of the effective number of HIV-1 B_{CAR} infections (*Ne*, *y* axis) along time (years, *x* axis) in Hispaniola under the BSKL and BSKG models. The median *Ne* estimates provided by the logistic growth (LG) parametric model (dark gray line) and their 95% HPD (pale gray area) are co-plotted in both graphics. **(C)** Plot summarizing the number of new HIV cases in adult (>15 years old) populations from Haiti and the Dominican Republic according to the UNAIDS estimations (http://aidsinfo.unaids.org/). The dashed vertical lines indicate the time of the last coalescent event.

DISCUSSION

This study confirms that the HIV-1 subtype B epidemic in Haiti is mostly driven by dissemination of early divergent non-pandemic B_{CAR} lineages (Cabello et al., 2014). The relative prevalence of B_{CAR} lineages among HIV-1 subtype B infections in Haiti (72%) closely matches that estimated in the neighboring Dominican Republic (74%) (Cabello et al., 2014). Together, Haiti and the Dominican Republic host the largest population (~160,000 people) of B_{CAR}-infected individuals in the Americas. Our results reveal a high degree of phylogenetic intermixing of the HIV-1 B_{CAR} sequences from Haiti and the Dominican Republic, consistent with the historical intense cross-border population mobility (Cohen, 2006; Rojas et al., 2011), but also support some level of geographic structure within Hispaniola. We estimate that nearly 55% of B_{CAR} infections in the Dominican Republic probably occurred within local transmission networks.

Previous evolutionary analyses (Gilbert et al., 2007; Junqueira et al., 2011; Worobey et al., 2016) support that the HIV-1 subtype B arrived in Haiti earlier than in any other American country. Here, we tested the hypothesis of the Haitian origin of subtype B epidemic by using for the first time a geographically balanced HIV-1 B_{CAR} dataset of Haitian (n = 103) and Dominican (n = 123) sequences. Our Bayesian phylogeographic analysis traced the root location of the B_{CAR} epidemic into Haiti and the Dominican Republic with the same posterior probability. This result revealed the complexity to uncover the geographic origin of the subtype B epidemic within Hispaniola by using only genetic data, probably due to the continuous cross-border viral movements between countries since the very early epidemic times. Thus, integration of non-genetic information (such as HIV incidence and prevalence data and human flows) into phylogeographic inference (Graf et al., 2015) would be probably indispensable to resolve the precise location of the HIV-1 B_{CAR} ancestor within Hispaniola.

Although phylogeographic analyses conducted here support the Haitian or Dominican origin of subtype B with the same probability, some epidemiological and historical data favor the Haitian origin hypothesis. The first AIDS cases were recognized in Haiti (1978–1979) a few years before those of the Dominican Republic (1983) and, by the early 1980s, HIV seroprevalence among Haitians was higher than among Dominicans (Pape et al., 1983; Koenig et al., 1987). Furthermore, the estimated T_{MRCA} of the B_{CAR} ancestor at around the late 1960s coincides with the return of Haitian professionals from the Democratic Republic of Congo (Gilbert et al., 2007), a country with an established HIV epidemic by that time (Worobey et al., 2008). By contrast, we found no similar historical link supporting relevant human flows between the Dominican Republic and the Democratic Republic of Congo during the 1960s.

Some epidemiological studies suggested that HIV-1 transmission between populations in Haiti and the Dominican Republic was uncommon during the early years and that tourists were the most likely source of first virus transmissions to Dominicans (Pape et al., 1983; Koenig et al., 1987). Our phylogeographic analysis, however, supports that HIV-1 B_{CAR} strains have been disseminated between Haiti and the Dominican Republic since the early 1970s and that several B_{CAR} transmission networks were already established in the Dominican Republic by mid-1970s, nearly a decade before the first AIDS cases were officially recognized in the country. The overall time scale here obtained for the B_{CAR} epidemic is fully consistent with that recovered in previous studies (Gilbert et al., 2007; Junqueira et al., 2011; Cabello et al., 2014; Worobey et al., 2016), supporting the reliability of our T_{MRCA} estimates. These results clearly indicate that HIV-1 B_{CAR} strains have been disseminated between Haitian and Dominican populations quite frequently since the early 1970s onwards.

While the BSKL model supports a stabilization of the B_{CAR} epidemic in Hispaniola from the early 1990s onwards, the BSKG reconstruction supports a sustained expansion until the late 1990s and a subsequent decline until the most recent coalescent event. The pattern here recovered by the BSKG model is consistent with the reported decline in HIV incidence in Haiti and the Dominican Republic since the late 1990s (UNAIDS, 2013, 2017), which is likely partially driven by changes in sexual behavior since the mid-1990s (Halperin et al., 2009). This finding is in agreement with previous studies that described that the BSKG model may uncover some aspects of the population history undetected by other Bayesian models (Gill et al., 2013; Mir et al., 2018). The mean growth rate estimated here for the B_{CAR} epidemic in Hispaniola during the first decades (0.50 $year^{-1}$) is similar to those estimated for B_{CAR} and $B_{PANDEMIC}$ lineages spreading in American countries with generalized heterosexual epidemics (0.35–0.45 $year^{-1}$) (Cabello et al., 2014; Mendoza et al., 2014; Mir et al., 2015; Bello et al., 2018).

A drawback to consider in our study is the relative small size of our sample. According to the UNAIDS, the number of people living with HIV was estimated at around 150,000 in Haiti and 67,000 in the Dominican Republic in 2017. Assuming that 70–75% of those infections probably correspond to B_{CAR} viruses, a very small fraction (<1%) of B_{CAR}-infected people living in those countries was included in our study. This low sampling density does not provide adequate power to assess HIV-1 clustering in generalized epidemics and could produce misleading results (Novitsky et al., 2014). The second limitation of our study is the lack of meta-data (such as city of origin, age, sex, or potential risk behavior) for most of the HIV-infected individuals included in our analysis, avoiding the identification of trends between individuals linked within the same local cluster. Finally, it is unclear whether our sample truly represents the whole diversity of HIV-1 B_{CAR} in Haiti because most Haitian individuals here analyzed were from the capital city (Port-au-Prince).

In summary, this study highlights that the HIV-1 epidemic in Haiti is mainly driven by dissemination of early divergent non-pandemic B_{CAR} strains. Our findings revealed that the HIV-1 B_{CAR} epidemics in Haiti and the Dominican Republic are highly connected by intensive bidirectional viral dispersal since the early 1970s and that local B_{CAR} transmission was already established in both countries when the first AIDS cases were officially recognized. Despite the use of a geographically balanced B_{CAR} dataset, probabilistic Bayesian phylogeographic models cannot uncover the root of the subtype B epidemic in Hispaniola by using only genetic sequence information. Our

findings support that both national and bi-national coordinated prevention measures are necessary to further control the HIV-1 dissemination in Haiti and the Dominican Republic.

DATA AVAILABILITY

All HIV-1 sequences were deposited in the GenBank database (accession numbers MK639799–MK639925).

ETHICS STATEMENT

Comité National d'Ethique du Ministère de la Santé Publique et de la Population de Haiti (13/07/2009).

AUTHOR CONTRIBUTIONS

GS, GB, VL, and MN conceived and designed the study. GS, MO, JB, RC, and BL collected the samples and performed the HIV sequence amplification and genotyping. GB and IA performed the phylogenetic and phylodynamics inferences. MN performed the statistical analyses. GS, GB, IA, and VL wrote the manuscript. All authors analyzed the data and discussed and reviewed the manuscript.

FUNDING

IA was funded by a fellowship from "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil)". GB was funded by a fellowship from "Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq".

ACKNOWLEDGMENTS

We thank T. Graf for their support with phylogeographic analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01340/full#supplementary-material

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Baele, G., Lemey, P., and Suchard, M. A. (2016). Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Syst. Biol.* 65, 250–264. doi: 10.1093/sysbio/syv083

Bello, G., Nacher, M., Divino, F., Darcissac, E., Mir, D., and Lacoste, V. (2018). The HIV-1 subtype B epidemic in French Guiana and Suriname is driven by ongoing transmissions of pandemic and non-pandemic lineages. *Front. Microbiol.* 9:1738. doi: 10.3389/fmicb.2018.01738

Cabello, M., Junqueira, D., and Bello, G. (2015). Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. *AIDS* 29, 483–492. doi: 10.1097/QAD.0000000000000552

Cabello, M., Mendoza, Y., and Bello, G. (2014). Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. *PLoS One* 9:e106045. doi: 10.1371/journal.pone.0106045

Chen, J. H., Wong, K. H., Chan, K. C., To, S. W., Chen, Z., and Yam, W. C. (2011). Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS One* 6:e25286. doi: 10.1371/journal.pone.0025286

Cohen, J. (2006). HIV/AIDS: Latin America & Caribbean. Dominican Republic: a sour taste on the sugar plantations. *Science* 313, 473–475. doi: 10.1126/science.313.5786.473

De Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., et al. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21, 3797–3800. doi: 10.1093/bioinformatics/bti607

Divino, F., De Lima Guerra Corado, A., Gomes Naveca, F., Stefani, M. M., and Bello, G. (2016). High prevalence and onward transmission of non-pandemic HIV-1 Subtype B clades in northern and northeastern Brazilian regions. *PLoS One* 11:e0162112. doi: 10.1371/journal.pone.0162112

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

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307–1320.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103

Ferreira, M. A. R., and Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* 26, 355–368. doi: 10.1002/cjs.5550360302

Figueroa, J. P. (2008). The HIV epidemic in the Caribbean: meeting the challenges of achieving universal access to prevention, treatment and care. *West Indian Med. J.* 57, 195–203.

Gilbert, M. T., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E., and Worobey, M. (2007). The emergence of HIV/AIDS in the Americas and beyond. *Proc. Natl. Acad. Sci. USA* 104, 18566–18570. doi: 10.1073/pnas.0705329104

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., and Suchard, M. A. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* 30, 713–724. doi: 10.1093/molbev/mss265

Graf, T., Vrancken, B., Maletich Junqueira, D., De Medeiros, R. M., Suchard, M. A., Lemey, P., et al. (2015). Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *J. Virol.* 89, 12341–12348. doi: 10.1128/JVI.01681-15

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and 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

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–W559. doi: 10.1093/nar/gki352

Halperin, D. T., De Moya, E. A., Perez-Then, E., Pappas, G., and Garcia Calleja, J. M. (2009). Understanding the HIV epidemic in the Dominican Republic: a prevention success story in the Caribbean? *J. Acquir. Immune Defic. Syndr.* 51(Suppl. 1), S52–S59. doi: 10.1097/QAI.0b013e3181a267e4

Hue, S., Pillay, D., Clewley, J. P., and Pybus, O. G. (2005). Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc. Natl. Acad. Sci. USA* 102, 4425–4429. doi: 10.1073/pnas.0407534102

Junqueira, D. M., De Medeiros, R. M., Matte, M. C., Araujo, L. A., Chies, J. A., Ashton-Prolla, P., et al. (2011). Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One* 6:e27489. doi: 10.1371/journal.pone.0027489

Koenig, R. E., Pittaluga, J., Bogart, M., Castro, M., Nunez, F., Vilorio, I., et al. (1987). Prevalence of antibodies to the human immunodeficiency virus in Dominicans and Haitians in the Dominican Republic. *JAMA* 257, 631–634. doi: 10.1001/jama.1987.03390050057018

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/journal.pcbi.1000520

Lopez, P., Rivera-Amill, V., Paulino-Ramirez, R., and Yamamura, Y. (2015). Short communication: HIV-1 subtype B in the dominican republic: evolution and molecular epidemiology. *AIDS Res. Hum. Retrovir.* 31, 679–684. doi: 10.1089/AID.2014.0304

Martin, D. P., Posada, D., Crandall, K. A., and Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retrovir.* 21, 98–102. doi: 10.1089/aid.2005.21.98

Mendoza, Y., Martinez, A. A., Castillo Mewa, J., Gonzalez, C., Garcia-Morales, C., Avila-Rios, S., et al. (2014). Human immunodeficiency virus type 1 (HIV-1) subtype B epidemic in Panama is mainly driven by dissemination of country-specific clades. *PLoS One* 9:e95360. doi: 10.1371/journal.pone.0095360

Mir, D., Cabello, M., Romero, H., and Bello, G. (2015). Phylodynamics of major HIV-1 subtype B pandemic clades circulating in Latin America. *AIDS* 29, 1863–1869. doi: 10.1097/QAD.0000000000000770

Mir, D., Graf, T., Esteves De Matos Almeida, S., Pinto, A. R., Delatorre, E., and Bello, G. (2018). Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches. *Sci. Rep.* 8:877. doi: 10.1038/s41598-018-26824-4

Myers, J. E., Taylor, B. S., Rojas Fermin, R. A., Reyes, E. V., Vaughan, C., Jose, L., et al. (2012). Transmitted drug resistance among antiretroviral-naive patients with established HIV type 1 infection in Santo Domingo, Dominican Republic and review of the Latin American and Caribbean literature. *AIDS Res. Hum. Retrovir.* 28, 667–674. doi: 10.1089/aid.2010.0355

Nadai, Y., Eyzaguirre, L. M., Sill, A., Cleghorn, F., Nolte, C., Charurat, M., et al. (2009). HIV-1 epidemic in the Caribbean is dominated by subtype B. *PLoS One* 4:e4814. doi: 10.1371/journal.pone.0004814

Novitsky, V., Moyo, S., Lei, Q., Degruttola, V., and Essex, M. (2014). Impact of sampling density on the extent of HIV clustering. *AIDS Res. Hum. Retrovir.* 30, 1226–1235. doi: 10.1089/aid.2014.0173

O'brien, J. D., Minin, V. N., and Suchard, M. A. (2009). Learning to count: robust estimates for labeled distances between molecular sequences. *Mol. Biol. Evol.* 26, 801–814. doi: 10.1093/molbev/msp003

Pape, J. W., Liautaud, B., Thomas, F., Mathurin, J. R., St Amand, M. M., Boncy, M., et al. (1983). Characteristics of the acquired immunodeficiency syndrome (AIDS) in Haiti. *N. Engl. J. Med.* 309, 945–950. doi: 10.1056/NEJM198310203091603

Parker, J., Rambaut, A., and Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* 8, 239–246. doi: 10.1016/j.meegid.2007.08.001

Rambaut, A. (2009). FigTree v1.4: Tree Figure Drawing Tool. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed January 2019).

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Carvalho, L. M., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vew007. doi: 10.1093/ve/vew007

Rojas, P., Malow, R., Ruffin, B., Rothe, E. M., and Rosenberg, R. (2011). The HIV/AIDS epidemic in the Dominican Republic: key contributing factors. *J. Int. Assoc. Physicians AIDS Care* 10, 306–315. doi: 10.1177/1545109710397770

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

UNAIDS (2013). Report on the global AIDS epidemic. Available at: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf (accessed January 2019).

UNAIDS (2017). Fact sheet—Latest global and regional statistics on the status of the AIDS epidemic. Available at: http://www.unaids.org/en/resources/documents/2017/UNAIDS_FactSheet (accessed January 2019).

Wang, T. H., Donaldson, Y. K., Brettle, R. P., Bell, J. E., and Simmonds, P. (2001). Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J. Virol.* 75, 11686–11699. doi: 10.1128/JVI.75.23.11686-11699.2001

Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., et al. (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 455, 661–664. doi: 10.1038/nature07390

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., et al. (2016). 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539, 98–101. doi: 10.1038/nature19827

Zehender, G., Ebranati, E., Lai, A., Santoro, M. M., Alteri, C., Giuliani, M., et al. (2010). Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J. Acquir. Immune Defic. Syndr.* 55, 156–160. doi: 10.1097/QAI.0b013e3181eb3002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership