

Viruses previously identified in Brazil as belonging to HIV-1 circulating recombinant form (CRF) 72_BF1 represent two closely related CRFs, one of which, designated CRF122_BF1, is also circulating in Spain.

Javier E. Cañada-García¹, Elena Delgado¹, Horacio Gil¹, Sonia Benito¹, Mónica Sánchez¹, Antonio Ocampo², Jorge Julio Cabrera^{3,4}, Celia Miralles², Elena García-Bodas¹, Ana Mariño⁵, Patricia Ordóñez⁶, María José Gude⁷, Carmen Ezpeleta⁸, Michael M Thomson^{1*}.

¹HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III. Majadahonda, Madrid, Spain.

²Department of Internal Medicine. Complejo Hospitalario Universitario de Vigo. Vigo, Pontevedra, Spain.

³Department of Microbiology, Complejo Hospitalario Universitario de Vigo. Vigo, Pontevedra, Spain.

⁴Microbiology and Infectology Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur). SERGAS-UVIGO. Vigo, Pontevedra, Spain.

⁵Infectious Diseases Unit, Complejo Hospitalario Universitario de Ferrol. Ferrol, A Coruña, Spain.

⁶Department of Microbiology. Complejo Hospitalario Universitario de Ferrol. Ferrol, A Coruña, Spain.

⁷Department of Microbiology, Hospital Universitario Lucus Augusti. Lugo, Spain.

⁸Department of Clinical Microbiology, Complejo Hospitalario de Navarra. Pamplona, Navarra, Spain.

***Corresponding author:**

Michael M. Thomson

mthomson@isciii.es

Abstract

Circulating recombinant forms (CRFs) are important components of the HIV-1 pandemic. Those derived from recombination between subtype B and subsubtype F1, with 18 reported, most of them of South American origin, are among the most diverse. Here we identify a HIV-1 BF1 recombinant cluster that is expanding in Spain, transmitted mainly via heterosexual contact, which, analyzed in near full-length genomes in 4 viruses, exhibited a coincident BF1 mosaic structure, with 12 breakpoints, that fully coincided with that of two viruses from Brazil (10BR_MG003 and 10BR_MG005), previously classified as CRF72_BF1. The three remaining Brazilian viruses previously identified as CRF72_BF1 (10BR_MG002, 10BR_MG004, and 10BR_MG008), exhibited mosaic structures highly similar, but not identical, to that of the Spanish viruses and to 10BR_MG003 and 10BR_MG005, with discrepant subtypes in two short genome segments, located in *pol* and *gp120^{env}*. Based on these results, we propose that the 5 viruses from Brazil previously identified as CRF72_BF1 actually belong to two closely related CRFs, one comprising 10BR_MG002, 10BR_MG004, and 10BR_MG008, which keep their CRF72_BF1 designation, and the other, designated CRF122_BF1, comprising 10BR_MG003, 10BR_MG005 and the viruses of the identified Spanish cluster. Three other BF1 recombinant genomes, 2 from Brazil and 1 from Italy, previously identified as unique recombinant forms, were classified as CRF72_BF1. CRF122_BF1, but not CRF72_BF1, was associated with protease L89M substitution, reported to contribute to antiretroviral drug resistance. Phylodynamic analyses estimate the emergence of CRF122_BF1 in Brazil around 1987. Given their close phylogenetic relationship and similar structures, the grouping of CRF72_BF1 and CRF122_BF1 in a CRF family is proposed.

1. Introduction

HIV-1 is characterized by high genetic diversity and rapid evolution, derived from elevated mutation and recombination rates. Through these mechanisms, the HIV-1 group M, the causative agent of the AIDS pandemic, has evolved into numerous circulating genetic forms, known as subtypes, of which 10 have been identified (A-D, F-H, K, L), subsubtypes (A1-A6, F1, F2), and circulating recombinant forms (CRFs), 118 of which are currently listed in the Los Alamos HIV Sequence Database (HIV Sequence Database, 2021). Additionally, geographic variants and clusters, some representing substantial proportions of viruses in certain areas, have been identified through phylogenetic analyses within subtypes, subsubtypes, and CRFs (Thomson and Nájera, 2005; Delgado et al., 2015; Delgado et al., 2019). Genetic characterization of HIV-1 variants is of public health relevance, as it allows to track their geographic spread and to estimate their population growth and the efficacy of preventive interventions (Paraskevis et al., 2016; Rife et al., 2017; Vasylyeva et al., 2020). It has also biological and clinical relevance, as different biological properties have been associated with some HIV-1 variants (Kiwauka et al. 2008; Pérez-Álvarez et al., 2014; Kourí et al., 2015; Venner et al., 2016; Cid-Silva et al., 2018; Song et al., 2019).

The number of CRFs is increasing incessantly, due to both the continuous generation of recombinant forms where diverse HIV-1 variants meet in the same population (Nájera et al., 2002), some of which become circulating through introduction into transmission networks, and the identification of old previously undocumented CRFs. The proportion of CRFs in the HIV-1 pandemic has increased over time, representing around 17% infections in 2010-2015 (Hemelaar et al., 2020). Among CRFs, those derived from subtype B and subsubtype F1 are among the most numerous, 18 of which have been reported in the literature, most of them originated in South America. The most widely circulating CRF from South America is CRF12_BF, which circulates at high prevalences in Argentina and Uruguay, where CRFs related to CRF12_BF are frequently found (Thomson et al., 2000; Thomson et al., 2002; Carr et al., 2001). Four other CRF_BFs (numbers 17, 38, 44, and 89) related to CRF12_BF, as evidenced by shared breakpoints and phylogenetic

clustering, were subsequently identified in different South American countries (Aulicino et al., 2012; Ruchansky et al., 2009; Delgado et al., 2010; Delgado et al., 2021). Due to their common ancestry and similar structures, these 5 CRFs and related URFs have been proposed to constitute a ‘family’ of recombinant viruses (Thomson and Nájera, 2005; Zhang et al., 2010; Delgado et al., 2021). By contrast, Brazilian CRF_BFs (De Sá Filho et al., 2006; Guimaraes et al., 2008; Sanabani et al., 2010; Pessoa et al., 2014; Reis et al., 2017; Reis et al., 2019) and CRF66_BF (the latter found mainly in Paraguay and Paraguayans living in Spain) (Bacqué et al., 2021) are unrelated to CRF12_BF. Similarly to the viruses of the CRF12_BF family, close relations have been reported between some Brazilian CRF_BFs: CRF28_BF and CRF29_BF (De Sá Filho et al., 2008); and CRF70_BF and CRF71_BF (Pessôa et al., 2014a).

Here we report the spread of a BF1 cluster in Spain whose viruses exhibit a mosaic structure identical to two Brazilian viruses previously identified as CRF72_BF1 (Pessôa et al., 2014b; Pessôa et al., 2016), which would represent a new CRF, with the three other viruses classified as CRF72_BF1 showing highly similar, but not identical, structures. We propose that viruses previously identified as CRF72_BF1 actually belong to two closely related CRFs that constitute a CRF family.

2. Materials and methods

2.1. Samples

Plasma and whole blood samples from HIV-1-infected individuals were collected in Spain for antiretroviral drug resistance tests and for a molecular epidemiological study. The study was approved by the Committee of Research Ethics of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain. It did not require written informed consent by the study participants, as it used samples and data collected as part of routine clinical practice and patients’ data were anonymized without retaining data allowing individual identification.

2.2. PCR amplification and sequencing

An ~1.4 kb pol fragment in protease-reverse transcriptase (Pr-RT) was amplified from plasma-extracted RNA or from whole blood-extracted DNA by (RT-)PCR followed by nested PCR, as described previously (Delgado et al., 2015) and sequenced with the Sanger method using a capillary automated sequencer. Near full-length genome (NFLG) sequences were obtained for selected samples by RT-PCR/nested PCR amplification from plasma RNA in five overlapping segments and sequenced by the Sanger method, as described (Delgado et al., 2002; Sierra et al., 2005; Cañada et al., 2021). Newly derived sequences are deposited in GenBank under accessions OL982311-OL982317 and OL982320-OL982323.

2.3. Phylogenetic sequence analyses

Sequences were aligned with MAFFT v7 (Kato and Standley, 2013). Initial phylogenetic trees with all Pr-RT sequences obtained by us were constructed via approximate maximum likelihood with FastTree v2.1.10 (Price et al., 2010), using the general time reversible evolutionary model with CAT approximation to account for among-site rate heterogeneity, with assessment of node support with Shimodaira-Hasegawa (SH)-like local support values (Guindon et al., 2010). Subsequent maximum likelihood (ML) trees with sequences of interest were constructed with W-IQ-Tree (Trifinopoulos et al., 2016), using the best-fit substitution model selected by ModelFinder program (Kalyaanamoorthy et al., 2017), with assessment of node support with the ultrafast bootstrap approximation approach (Hoang et al., 2018). Trees were visualized with MEGA v7.0 (Kumar et al., 2016).

Mosaic structures were analyzed by bootscanning (Salminen et al., 1995) with SimPlot v1.3.5 (Lole et al., 1999). In these analyses, trees were constructed using the neighbor-joining method with the Kimura 2-parameter model and a window width of 250 nucleotides. The subtype affiliations of recombinant segments identified with SimPlot, whose breakpoints were more precisely located in the midpoint of transitions between B-F1 subtype-discriminating nucleotides (here defined as those differing between the 75% consensus sequences of subtype B and the Brazilian F1 strain), were further phylogenetically analyzed via ML and Bayesian inference.

These analyses were performed with IQ-Tree; PhyML (Guindon et al., 2010), using the best-fit evolutionary model selected by the SMS program (Lefort et al., 2017) and node support assessment with the approximate likelihood ratio test, SH-like procedure; and MrBayes v3.2 (Ronquist et al., 2012), using the GTR+G+I substitution model, with two simultaneous independent runs and 8 chains 2–5 million generations long, ensuring that both runs reached convergence, as determined by an average standard deviation of split frequencies < 0.01, discarding the first 50% of the trees in the posterior distribution as burn-in. For these analyses, we used as outgroup a reconstructed B-F1 ancestral sequence, considering the phylogenetic relationship between B and F subtypes (Zhu et al., 1998), obtained with IQ-Tree. The use of a reconstructed ancestral sequence as outgroup is similar to the approach used in other studies (Thomson and Fernández-García, 2011; Travers et al., 2004; Seager et al., 2014), to prevent the long branch attraction artifact (Bergsten, 2005) that could be caused by an outgroup whose distance to the ingroup is relatively long compared to the within-ingroup distances. This artifact can result in collapse or substantial decrease in node support of the clades of the ingroup, particularly in short genome segments.

2.4. Antiretroviral drug resistance determination

Antiretroviral (ARV) drug resistance was analyzed with the HIVdb program of the Stanford University's HIV Drug Resistance Database (Rhee et al., 2003; Shafer, 2006).

2.5. Temporal and geographical estimations of clade ancestors

The time and the most probable location of the most recent common ancestor (MRCA) of the newly defined CRF were estimated using Pr-RT sequences with the Bayesian Markov chain Monte Carlo (MCMC) coalescent method implemented in BEAST v1.10.4 (Suchard et al., 2018). Before the BEAST analysis, the existence of temporal signal in the dataset was assessed with TempEst v1.5.3 (Rambaut et al., 2016). The BEAST analysis was performed using the SRD06 codon-based evolutionary model (where the 3rd codon position is in a partition different from the 1st and 2nd positions) (Shapiro et al., 2006). We also specified an uncorrelated

lognormal relaxed clock and a Bayesian SkyGrid coalescent tree prior (Gill et al., 2013). The MCMCs were run for 20 million generations. The runs were performed in duplicate, and the posterior tree files were combined with LogCombiner v1.10.4. Proper mixing of the chains was assessed with Tracer v1.6, ensuring that effective sample size values of all parameters were >200. The posterior distribution of trees was summarized in a maximum clade credibility (MCC) tree with TreeAnnotator v1.10.4, after removal a 10% burn-in. MCC trees were visualized with FigTree v1.4.2 (Rambaut, <http://tree.bio.ed.ac.uk/software/figtree/>). Parameter uncertainty was summarized in 95% highest posterior density (HPD) intervals.

3. Results

3.1. Identification of a HIV-1 cluster of F1 subtype in Pr-RT propagating in Spain

In a molecular epidemiological study in Spain, based on Pr-RT sequences, we detected frequent grouping in clusters, several of which were of F1 subtype in Pr-RT (Delgado et al., 2015; Delgado et al., 2019). One of them, designated F1_2, which is the focus of the present study, comprised 14 individuals, 13 of them from the region of Galicia, Northwest Spain (Table 1). Years of HIV-1 diagnoses were between 2007 and 2019 and transmission was predominantly heterosexual. Most individuals were Spanish but 3 were Brazilian and 1 was Ukrainian. To determine whether other sequences from databases belonged to this cluster, we performed BLAST searches in the HIV Sequence Database (Los Alamos National Laboratory, 2021), incorporating the most similar sequences in the phylogenetic analyses. This allowed to identify 3 additional sequences that belonged to the F1_2 cluster, from Brazil, Portugal, and Germany respectively (Figure 1). All but two of the viruses collected in Spain and the virus from Germany branched in a subcluster. Viruses from the F1_2 cluster were most closely related to viruses of the Brazilian F1 strain and to Brazilian CRF_BFs with Pr-RT derived from it.

Table 1. Epidemiological data of patients and GenBank accessions of sequences.

Sample ID	City of sample collection	Region of sample collection	Country of origin	Year of HIV diagnosis	Year of sample collection	Gender	Transmission route*	PR-RT GenBank accession	NFLG GenBank accession
X2592	Vigo	Galicia	Spain	2008	2008	F	HT	GU326146	-
X2632	Ferrol	Galicia	Spain	2009	2009	M	MSM	GU326158	KC113006 JX140660
X2657	Vigo	Galicia	Spain	2009	2009	F	HT	GU326163	-
GA076319	Vigo	Galicia	Ukraine	2010	2010	F	HT	OL982311	-
GA099170	Vigo	Galicia	Switzerland	2011	2011	F	HT	-	OL982312
GA330265	Vigo	Galicia	Brazil	2007	2007	Trans	Sexual	OL982313	-
GA486085	Ferrol	Galicia	Spain	2018	2018	F	n.a.	-	OL982314
GA501952	Vigo	Galicia	Spain	2014	2014	F	HT	OL982315	-
GA513250	Vigo	Galicia	Spain	2012	2012	M	Sexual	OL982316	-
GA522821	Lugo	Galicia	Brazil	2019	2019	M	MSM	OL982317	-
GA817166	Vigo	Galicia	Spain	2008	2016	F	HT	OL982320	-
GA874035	Vigo	Galicia	Spain	2012	2012	M	HT	-	OL982321
GA903064	Vigo	Galicia	Spain	2012	2012	F	HT	OL982322	-
NA584314	Pamplona	Navarre	Brazil	2017	2017	M	MSM	OL982323	-

*n.a.: datum not available; Trans: transgender; HT: heterosexual; MSM: man who has sex with men; sexual: unspecified sexual.

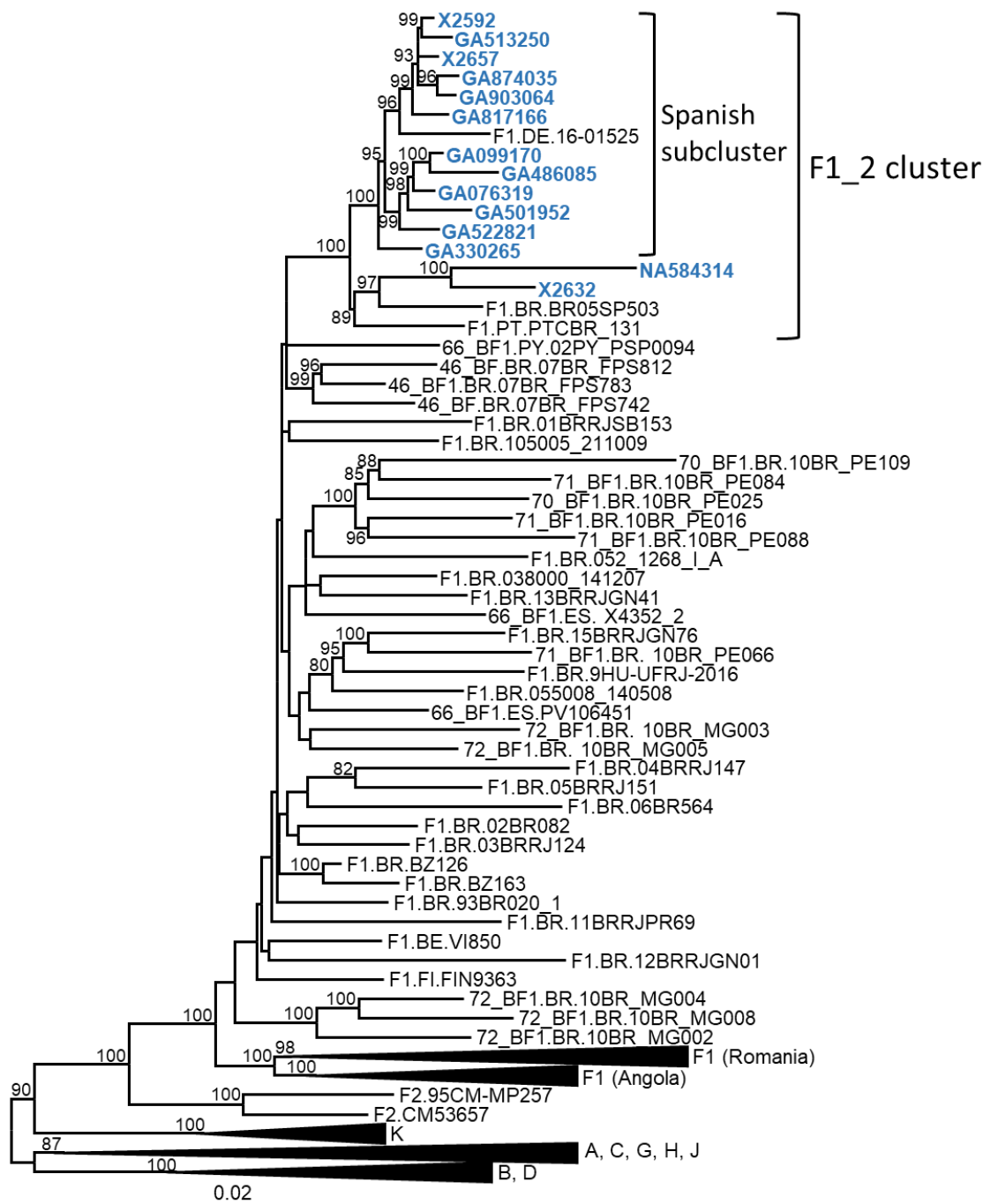


Figure 1. Maximum likelihood phylogenetic tree of Pr-RT sequences of the F1_2 cluster.

Names of sequences obtained by us, all collected in Spain, are in blue. Only ultrafast bootstrap values $\geq 80\%$ are shown. In database sequences, the country of sample collection is indicated before the virus name with the 2-letter ISO country code. BE: Belgium; BR: Brazil; DE: Germany; ES: Spain; FI: Finland; PT: Portugal; PY: Paraguay.

3.2. Analyses of NFLG sequences

In order to determine whether the F1_2 cluster was of uniform subtype or recombinant, we obtained NFLG sequences from 3 individuals from 2 cities through amplification from plasma RNA. A fourth NFLG sequence had been obtained previously from virus culture supernatant (Sanchez et al., 2014). Preliminary analyses of the NFLG with Recombination Identification Program (<https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>) indicated that the genomes were BF1 recombinant. To determine whether they belonged to a known CRF, we constructed a phylogenetic tree in which genomes of all CRF_BFs were included. The tree showed that viruses of the F1_2 cluster grouped in a strongly supported clade with viruses classified as CRF72_BF1, with two of them, 10BR_MG003 and 10BR_MG005, being the most closely related to the viruses of the F1_2 cluster, and the 3 others, 10BR_MG002, 10BR_MG004, and 10BR_MG008, branching in a sister clade (Figure 2).

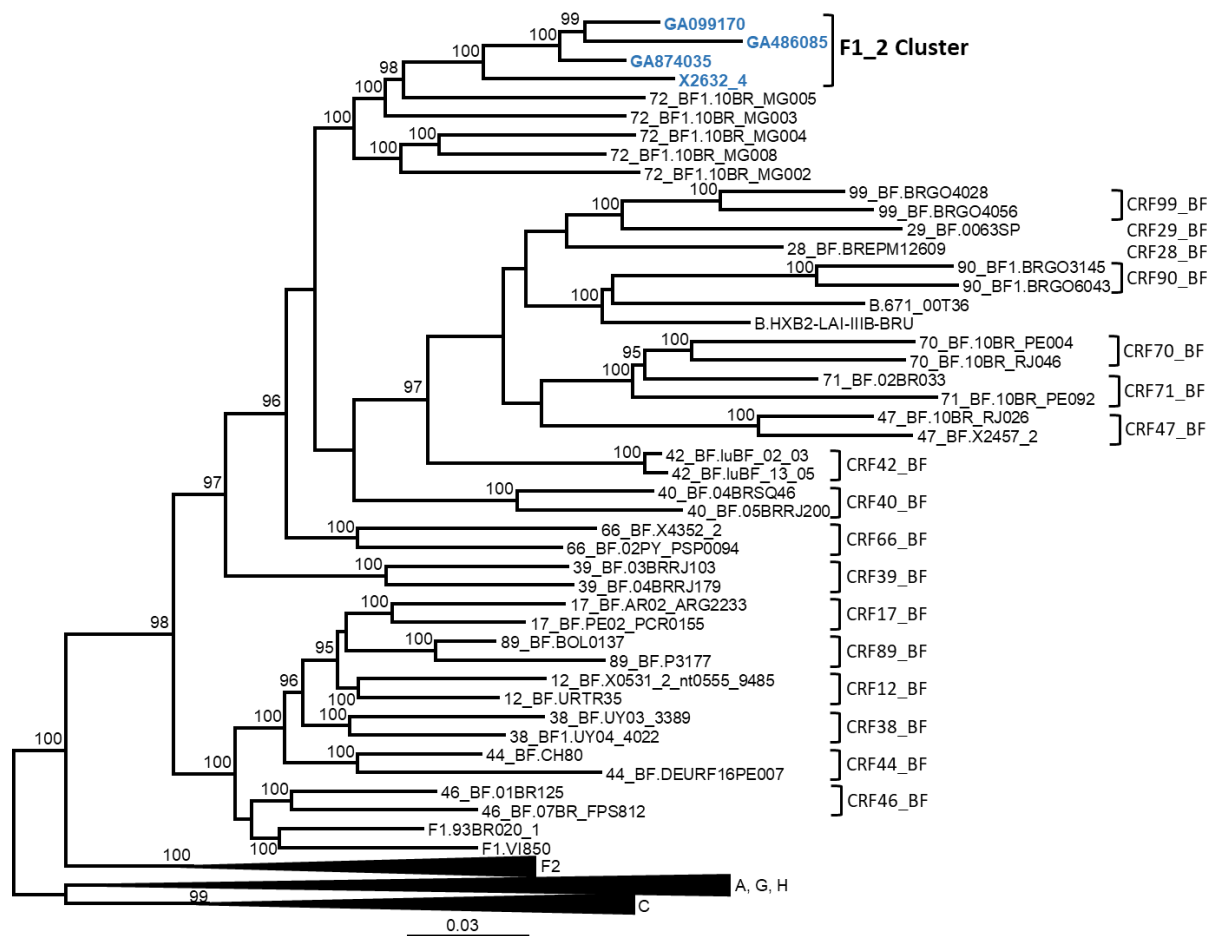


Figure 2. Maximum likelihood tree of NFLG sequences of viruses of the F1_2 cluster.

References of published CRF_BFs and of HIV-1 subtypes are also included in the analysis. Names of sequences obtained by us are in blue. In reference sequences, the subtype or CRF, as classified in the Los Alamos HIV Sequence Database, is indicated before the virus name. Only ultrafast bootstrap values $\geq 90\%$ are shown.

Bootscan analyses of NFLG sequences showed that the viruses of the F1_2 cluster were BF1 recombinant, exhibiting mosaic structures fully coincident with those of 10BR_MG003 and 10BR_MG005, and slightly different from the three other viruses classified as CRF72_BF1 (Figure 3 and supplementary Figure 1). The differences with these three viruses were observed in a short *pol* segment, around protease-reverse transcriptase (RT) junction, where grouping with subtype references was discrepant, and in the 5' segment of gp120, where the location of a BF1 breakpoint differed. The mosaic structures determined with bootscanning were confirmed by ML and Bayesian phylogenetic analyses of partial genome segments, which confirmed the coincidence of the mosaic structures of the 4 F1_2 viruses and the Brazilian 10BR_MG003 and 10BR_MG005 viruses and the subtype discrepancy in two genome segments (HXB2 positions 2429-2618 and 6432-6519) of these viruses with 10BR_MG002, 10BR_MG004, and 10BR_MG008 (Figure 4). These analyses, therefore, allowed to determine that viruses of the identified Spanish BF1 cluster, together with the Brazilian viruses 10BR_MG003 and 10BR_MG005, previously classified as CRF72_BF1, belong to a CRF, which was designated CRF122_BF1, which is closely related to, but different from, the three other viruses previously classified as CRF72_BF1, 10BR_MG002, 10BR_MG004, and 10BR_MG008, whose original CRF designation is maintained. The mosaic structures of both CRFs, as inferred from bootscan analyses, ML and Bayesian phylogenetic trees of partial sequences, and examination of intersubtype transitions of subtype-discriminating nucleotides, are shown in Figure 5.

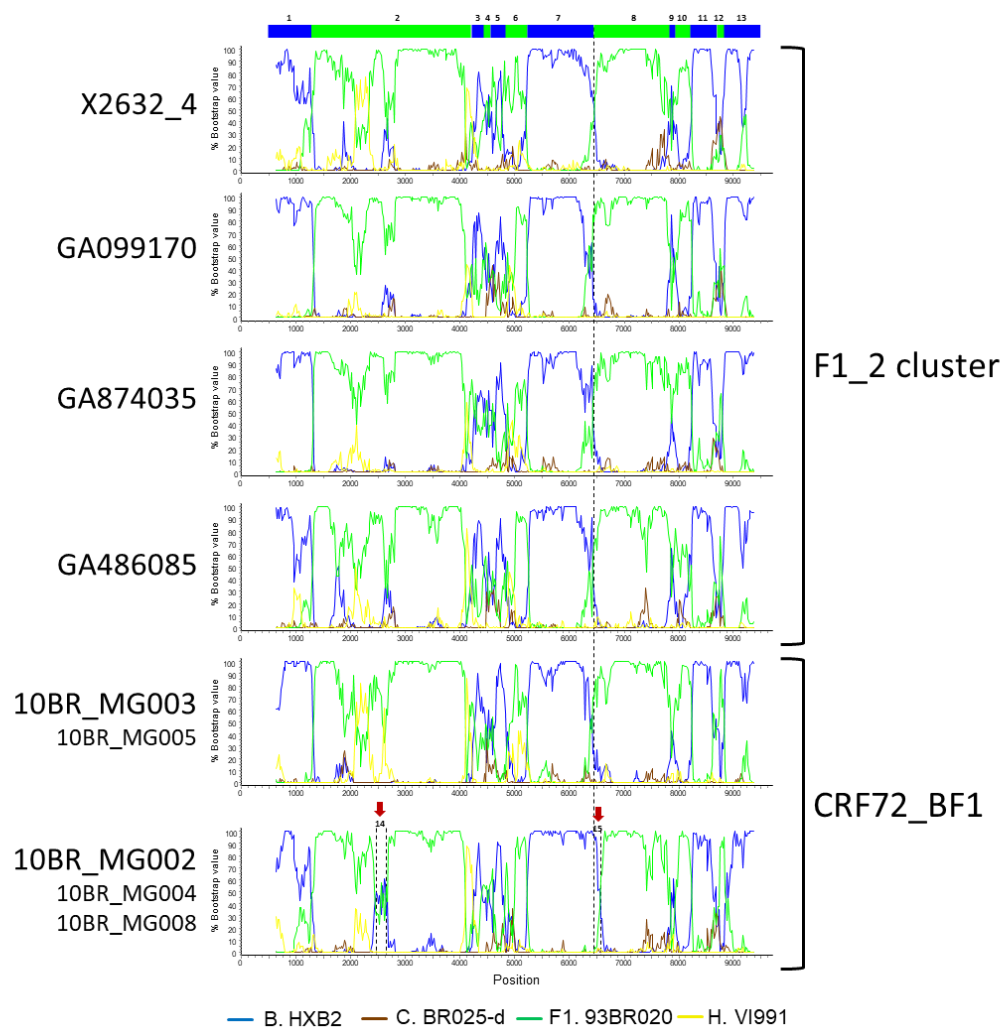


Figure 3. Bootscan analyses of NFLG sequences of viruses of the F1_2 cluster compared to those of viruses previously classified as CRF72_BF1. Bootscan plots of all four F1_2 viruses are shown, together with those of 10BR_MG003 and 10BR_MG002. The bootscan plot of 10BR_MG005 is almost identical to that of 10BR_MG003 and those of 10BR_MG004 and 10BR_MG008 are almost identical to that of 10BR_MG002, and are shown in Supplementary Figure 1. The horizontal axis represents the position in the HXB2 genome of the midpoint of a 250 nt window moving in 20 nt increments and the vertical axis represents bootstrap values supporting clustering with subtype reference sequences. The vertical dashed lines indicate B-F1 breakpoints. The bar on top indicates the numbers of the segments that were further analyzed with ML and Bayesian trees (Figure 4). Two genome segments that appear to group with different subtypes in 10BR_MG002, 10BR_MG004, and 10BR_MG008 relative to 10BR_MG003, 10BR_MG005, and the F1_2 viruses, are signaled with arrows above the bootscan plot of 10BR_MG002 and were also analyzed via ML and Bayesian inference (Figure 4).

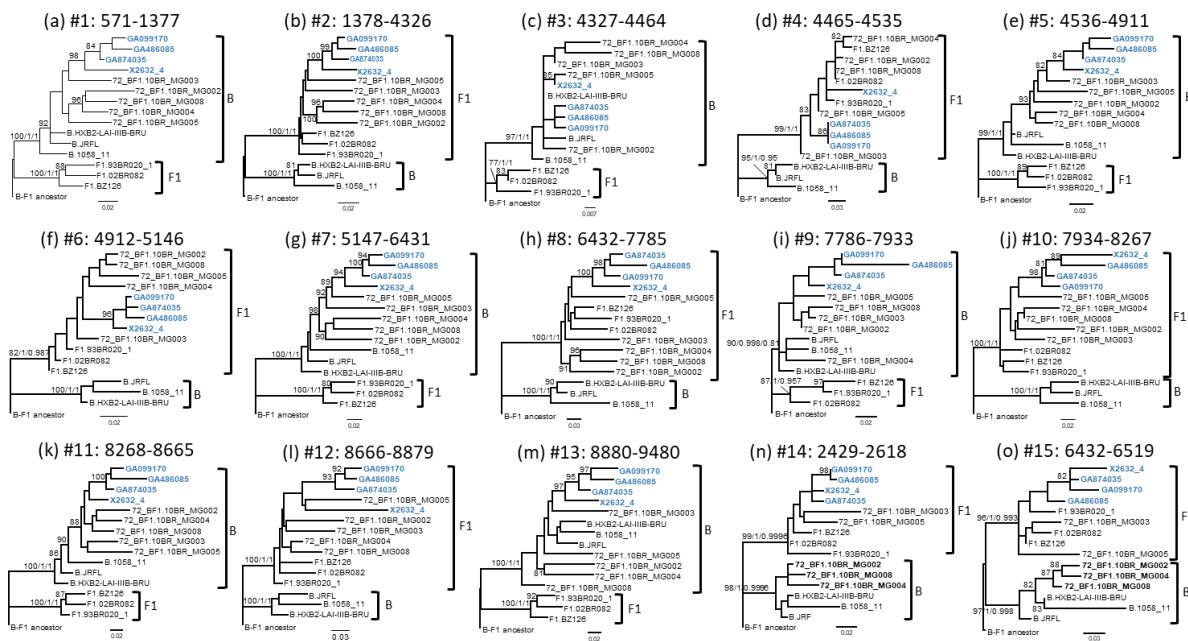


Figure 4. Phylogenetic trees of interbreakpoint genome segments of F1_2 viruses.

Breakpoints were defined according to the bootscan analyses and to midpoints of transitions between subtype-discriminating nucleotides, here defined as those where the 75% consensus of subtype B and of the Brazilian variant of subtype F1 differ. HXB2 positions delimiting the analyzed segments and their numbers as indicated in Figure 3 are indicated on top of the trees. Sequence names of F1_2 viruses are in blue. Names of subtype reference sequences are preceded by the corresponding subtype name. Sequences of viruses previously classified as CRF72_BF1 were also included. The trees are rooted with a reconstructed B-F1 ancestor. Node supports for B and F1 clades are indicated, in this order, as ultrafast bootstrap value/aLRT SH-like support/posterior probability, which were obtained with IQ-Tree, PhyML, and MrBayes programs, respectively. For the other nodes, only ultrafast bootstrap values $\geq 80\%$ are indicated. Trees in (n) and (o) (segments 2429-2618 and 6432-6519, respectively) correspond to the segments indicated with arrows in Figure 3 where F1_2 viruses and 10BR_MG003 and 10BR_MG005, on the one hand, and 10BR_MG002, 10BR_MG004, and 10BR_MG008, on the other, appeared to differ in subtype affiliations in the bootscan analyses.

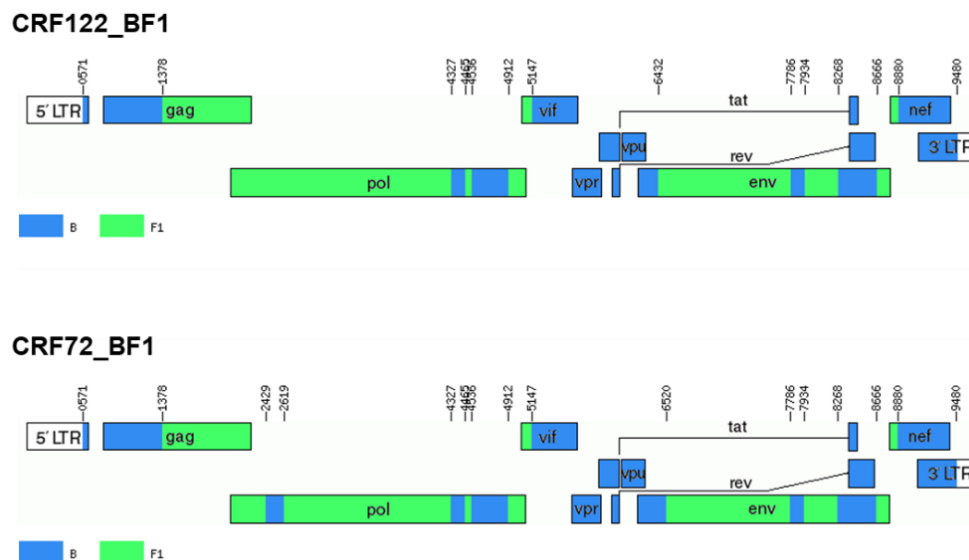


Figure 5. Mosaic structure of CRF72_BF1 and CRF122_BF1. Breakpoint positions are numbered as in the HXB2 genome. The drawing was made using the Recombinant HIV-1 Drawing Tool https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html.

Three additional BF1 recombinant NFLGs, originally identified as unique recombinant forms, two from Brazil, 99UFRJ-2 (Thomson et al., 2004) and BREPM1029 (Sa-Filho et al. 2007), and 1 from Italy, IT_BF_PRIN_454 (Bruselles et al., 2009), in their published analyses, exhibited mosaic structures similar to CRF72_BF1 and CRF122_BF1. To determine whether they belonged to one of these CRFs, we constructed a phylogenetic tree with NFLG sequences including the 3 mentioned genomes, which showed that all of them grouped with CRF72_BF1 viruses (Supplementary Figure 2). Bootscan analyses showed mosaic structures of 99UFRJ-2 and IT_BF_PRIN_454 coincident with that of CRF72_BF1; however the bootscan plot of BREPM1029 failed to show clustering with the subtype B reference in the protease-RT junction (HXB2 positions 2429-2618) (Supplementary Figure 3). Examination of subtype-discriminating nucleotides suggested that the 2429-2618 segment was of subtype B in 99UFRJ-2 and BREPM1029, which was confirmed in phylogenetic analyses of this fragment (Supplementary Figure 4a). However, in BREPM1029, the subtype B fragment in the protease-RT junction appeared to be slightly shorter, located between HXB2 positions 2479-2618, which was confirmed in phylogenetic trees (Supplementary Figure 4b). Phylogenetic analyses also showed that

in all three genomes the 6432-6519 segment in gp120 was of subtype B, as in CRF72_BF1 (Supplementary Figure 4c). These results allowed to confidently classify 99UFRJ-2 and IT_BF_PRIN_454 as CRF72_BF1 viruses. As to BREPM1029, given its strong phylogenetic clustering with CRF72_BF1 references and its minimal difference in mosaic structure with CRF72_BF1, with a breakpoint displaced only around 50 nt relative to this CRF, it seems reasonable to also classify it as CRF72_BF1, although we cannot definitively discern whether its breakpoint displacement is due to a different recombination event or to mutations occurring near the CRF72_BF1 breakpoint.

3.3. Differences in amino acid residues

We analyzed amino acid residues in viral proteins differing between CRF72_BF1 and CRF122_BF1 viruses and conserved within each CRF. We found 10 such amino acid residues (Table 2). One of them is in position 89 of protease, where CRF72_BF1 has leucine, which is the subtype B consensus, while CRF122_BF1 has methionine, which is the F1 subsubtype consensus. Protease L89M substitution has been reported to contribute, together with other protease mutations, to resistance to some protease inhibitor drugs (Calazans et al., 2005; Marcelin et al., 2008; Wensing et al., 2019).

Table 2. Differences in amino acid residues between CRF72_BF1 and CRF122_BF1*.

		p17 ^{gag}	PR		RT	IN	Vif	Tat	Rev	Vpu	gp120 ^{env}
		61	61	89	399	84	151	68	22	80	69
B	HXB2	L	Q	L	E	I	A	S	L	D	D
CRF72_BF	10BR_MG002	I	Q	L	E	L	V	S	L	D	D
	10BR_MG004	I	Q	L	E	L	A	S	L	D	D
	10BR_MG008	I	Q	L	D	L	A	S	L	D	D
	99UFRJ-2	I/M	Q	L	E	L	A	P	L	D	D
	BREPM1029	L	N	L	E	L	A	S	L	D	D
	IT_PRIN_454	I	Q	L	E	L	A	S	L	D	D
CRF122_BF1	10BR_MG003	L	N	M	D	I	T	D	I	N	N
	10BR_MG005	L	N	M	D	I	T	D	I	N	N
	X2632_4	L	N	M	D	L	T	D	I	N	N
	GA099170	L	N	M	D	I	T	D	I	N	N
	GA486085	L/I	N	M	D	I/M	T	D	I	N	D
	GA874035	L	N	M	D	I	T	D	I	N	N

*Only amino acid residues conserved in at least 5 of 6 CRF72_BF1 viruses and in both Brazilian and at least 3 of 4 Spanish CRF122_BF1 viruses are included in the table.

3.4. ARV drug resistance mutations

Primary ARV drug resistance mutations were detected in 2 CRF122_BF1 viruses, both from Brazil, 1 with L90M protease inhibitor (PI) resistance mutation and the other with D30N PI resistance mutation and M41L, D67N, M184V, and T215Y mutations of resistance to nucleoside RT inhibitors. .

3.5. Temporal and geographical estimation of CRF122_BF1 origin

To estimate the time and place of origin of CRF122_BF1, Pr-RT sequences were analyzed with the Bayesian coalescent method implemented in BEAST 1.10.4. Prior to this analysis, we performed TempEst analyses to determine whether there was an adequate temporal signal in the dataset. We found that the temporal signal, assessed by the correlation between root-to-tip distance and time, increased by masking the positions of codons with drug resistance mutations in any of the sequences ($r^2 = 0.5265$). Therefore, the BEAST analysis was performed with a sequence alignment where these codons had been removed. In this analysis, the substitution rate was estimated at 1.829×10^{-3} subs/site/year (95% HPD, $1.118 \times 10^{-3} - 2.542 \times 10^{-3}$ subs/site/year) and the time of the MRCA of CRF122_BF1 was estimated around 1987 (95% HPD, 1976-1998), with its most probable location being Brazil (location PP = 0.89) (Figure 6). The introduction of CRF122-BF1 in Spain (according to the MRCA of the Spanish cluster) was estimated in the Galician city of Vigo (location PP=0.992) around 2002 (95% HPD, 1998-2005).

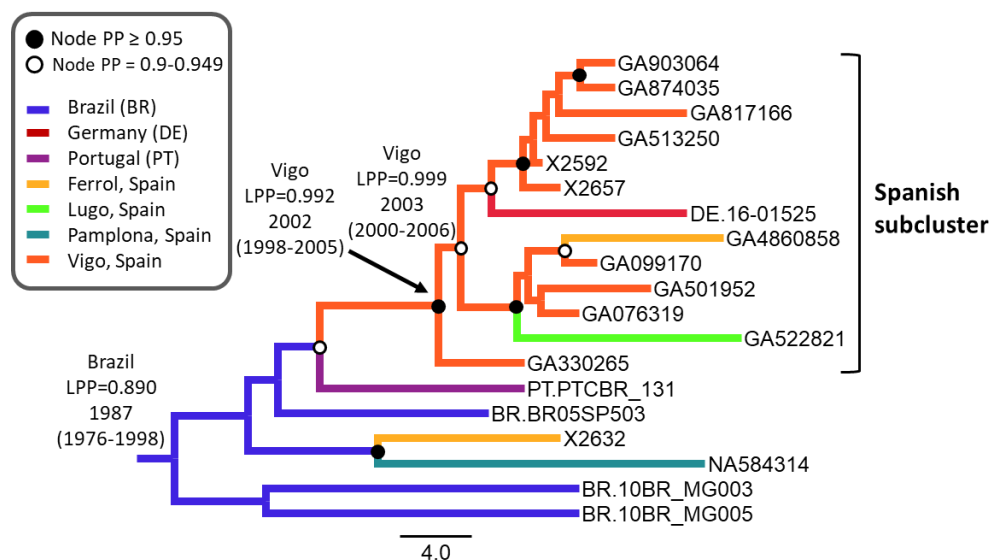


Figure 6. Maximum clade credibility tree of CRF122_BF Pr-RT sequences. Branch colors indicate, for terminal branches, country of sample, and for internal branches, the most probable location country of the subtending node, according to the legend on the upper left. Nodes supported by $PP \geq 0.95$ and $PP 0.9-0.949$ are indicated with filled and unfilled circles, respectively. The most probable locations at the root of the tree and at the node of the Spanish cluster are indicated, together with the PPs supporting each location (LPPs) and the year estimated for the MRCAs (mean value, with 95% HPD interval in parentheses).

4. Discussion

The results here presented indicate that the five Brazilian viruses previously classified as CRF72_BF1 actually belong to two closely related CRFs, one of which is circulating in Spain. Consequently, the CRF comprising two Brazilian viruses previously classified as CRF72_BF1 and the four Spanish viruses with coincident mosaic structures are given a new designation, CRF122_BF1, while the three other Brazilian viruses previously classified as CRF72_BF1 keep their original designation. Three other BF1 viruses analyzed in NFLGs originally classified as URFs, two from Brazil and one from Italy, were also classified on the basis of phylogenetic and bootscan analyses as CRF72_BF1. The close relationship between CRF122_BF1 and CRF72_BF1 is one more example of closely related CRFs, with precedents in South America. Other examples are CRF12_BF, CRF17_BF, and CRF89_BF (and more distantly related CRF38_BF and CRF44_BF) (Delgado et al., 2021); CRF28 and CRF29_BF (De Sá Filho et al., 2006); and CRF70_BF and CRF71_BF (Pessôa et al., 2014a).

Failure to realize that the five viruses previously identified as CRF72_BF1 represent two different CRFs may derive from the short segments in which both CRFs differ in subtypes. These differences may be missed if bootscan analyses are performed using window widths much greater than the length of the recombinant fragment. We have also noticed that jpHMM (Schultz et al., 2009), used in a previous study to analyze CRF72_BF1 genomes (Pessôa et al., 2006), often fails to detect short recombinant segments.

Given the close relationship and partial coincidence in mosaic structures of CRF72_BF1 and CRF122_BF1, we propose that they are members of a CRF family, similarly to the

CRF family of BF1 recombinant viruses from South America comprising CRFs numbers 12, 17, 38, 44, and 89 (Delgado et al., 2021). The grouping of some closely related HIV-1 recombinants derived from a common recombinant ancestor in families was proposed by us (Thomson and Nájera, 2005) and others (Zhang et al., 2010). Proposed families of CRFs with close phylogenetic relations, shared parental strains, and partially coincident breakpoints are indicated in the phylogenetic tree of Supplementary Figure 5.

The estimated origin of CRF122_BF1 around 1987 is consistent with the estimated origin of the Brazilian F1 strain (around 1977) (Bello et al., 2006) and similar to those of other South American CRF_BFs (CRF12, CRF28/29, CRF38, CRF89, and CRF90) reported in the literature (Bello et al., 2010; Ristic et al., 2011; Reis et al., 2017; Delgado et al., 2021), but younger than some other estimates for CRF12_BF in the 1970s (Dilernia et al., 2011; Delgado et al., 2021) and older than the estimates for CRF99_BF, around 1993 (Reis et al., 2017).

The correct classification of HIV-1 genetic forms is important, since even relatively minor genetic differences in viral genomes may result in important biological differences. Examples in HIV-1 are CXCR4 coreceptor usage in CRF14_BG, which is associated with only four amino acid residues in the Env V3 loop (Pérez-Álvarez et al., 2014), all or most of which are absent in viruses of the closely related CRF73_BG (Fernández-García et al., 2016), which has a very similar mosaic structure; and differences in pathogenic potential or therapeutic response associated with clusters within HIV-1 CRF01_AE (Song et al., 2019) and F1 subsubtype (Cid-Silva et al., 2018). Here we show that CRF122_BF1, but not CRF72_BF1, has the protease L89M substitution, that has been reported to contribute, together with other protease mutations, to resistance to tipranavir/ritonavir (Marcelin et al., 2008) and, within an F1 subsubtype background, to resistance to other protease inhibitor drugs (Calazans et al., 2005).

CRF122_BF1 represents one more example of a CRF of South American ancestry first identified in Western Europe. Others are CRF42_BF (Struck et al., 2015), CRF47_BF (Fernández-García et al., 2010), CRF60_BC (Simonetti et al., 2014), CRF66_BF (Bacqué et al., 2021), and CRF89_BF (Delgado et al., 2021). This may derive from the increasing migratory flows from South America to Europe and from the relatively low number of

HIV-1 sequences available in some South American countries (Bacqué et al., 2021). Therefore, HIV-1 molecular epidemiological studies in Europe may contribute to a better knowledge of the HIV-1 epidemics in South America.

In summary, we show that viruses of a BF1 recombinant cluster of Brazilian ancestry circulating in Spain exhibit a mosaic structure that is fully coincident with that of two Brazilian viruses previously classified as CRF72_BF1, and highly similar, but not identical, to that of three other Brazilian viruses also classified as CRF72_BF1. Therefore, we propose a new CRF designation, CRF122_BF1, for the viruses of the Spanish cluster and the two Brazilian viruses with coincident structures, which together with CRF72_BF1 would constitute a CRF family. The accurate genetic characterization of HIV-1 variants is important to determine their associated biological features and to track their epidemic spread.

Acknowledgments

We thank José Antonio Taboada, from Consellería de Sanidade, Xunta de Galicia, for his support of this study, and the personnel at the Genomic Unit, Instituto de Salud Carlos III, for technical assistance in sequencing.

Author contributions

MT and ED conceived the study and supervised the experimental work. JC-G, ED, and MT processed sequences and performed phylogenetic and phylodynamic analyses. HG performed data curation. JC-G, SB, MS and EG-B performed experimental work. AO, JJC, CM, AM, PO, MJG, and CE obtained samples and epidemiological data from patients. MT, JC-G, ED, and HG wrote the manuscript. All authors read and approved the manuscript.

Funding

This work was funded through Acción Estratégica en Salud Intramural (AESI), Instituto de Salud Carlos III, projects “Estudios sobre vigilancia epidemiológica molecular del VIH-1 en España”, PI16CIII/00033, and “Epidemiología molecular del VIH-1 en España y su

utilidad para investigaciones biológicas y en vacunas“, PI19CIII/00042, and through scientific agreement with Consellería de Sanidade, Government of Galicia (MVI 1004/16).

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.

References

1. Aulicino, P. C., Gómez-Carrillo, M., Bello, G., Rocco, C., Mangano, A., Carr, J., et al. (2012). Characterization of full-length HIV-1 CRF17_BF genomes and comparison to the prototype CRF12_BF strains. *Infect. Genet. Evol.* 12, 443-447. doi:10.1016/j.meegid.2012.01.003.
2. Bacqué, J., Delgado, E., Benito, S., Moreno-Lorenzo, M., Montero, V., Gil, H., et al. (2021). Identification of CRF66_BF, a new HIV-1 circulating recombinant form of South American origin. *Front. Microbiol.* 12, 774386. doi: 10.3389/fmicb.2021.774386.
3. Bello, G., Eyer-Silva, W. A., Couto-Fernandez, J. C., Guimarães, M. L., Chequer-Fernandez, S. L., Teixeira, 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.
4. Bello, G., Aulicino, P. C., Ruchansky, D., Guimarães, M. L., López-Galíndez, C., Casado, C., et al. (2010). Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. *Retrovirology* 7, 22. doi: 10.1186/1742-4690-7-22.
5. Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21:163-193. doi: 10.1111/j.1096-0031.2005.00059.x.
6. Bruselles, A., Rozera, G., Bartolini, B., Prospero, M., Del Nonno, F., Narciso, P., et al. (2009). Use of massive parallel pyrosequencing for near full-length characterization of a unique HIV Type 1 BF recombinant associated with a fatal primary infection. *AIDS Res. Hum. Retroviruses* 25, 937-942. doi:10.1089/aid.2009.0083.
7. Calazans, A., Brindeiro, R., Brindeiro, P., Verli, H., Arruda, M. B., Gonzalez, L. M., et al. (2005). Low accumulation of L90M in protease from subtype F HIV-1 with

- resistance to protease inhibitors is caused by the L89M polymorphism. *J. Infect. Dis.* 191, 1961-1970. doi:10.1086/430002.
8. Cañada, J. E., Delgado, E., Gil, H., Sánchez, M., Benito, S., García-Bodas, E., et al. (2021). Identification of a new HIV-1 BC intersubtype circulating recombinant form (CRF108_BC) in Spain. *Viruses* 13, 93. doi: 10.3390/v13010093.
 9. Carr, J. K., Avila, M., Gomez Carrillo, M., Salomon, H., Hierholzer, J., Watanaveeradej, V., et al. (2001). Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS* 15, F41-F47. doi:10.1097/00002030-200110190-00002.
 10. 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.
 11. De Sá Filho, D. J., Sucupira, M. C., Caseiro, M. M., Sabino, E. C., Diaz, R. S., and Janini, L. (2006). M. Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res. Hum. Retroviruses* 22, 1-13. doi: 10.1089/aid.2006.22.1.
 12. 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.
 13. 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.
 14. 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.

15. Delgado, E., Benito, S., Montero, V., Cuevas, M. T., Fernández-García, A., Sánchez-Martínez, M., et al. (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: <https://doi.org/10.3389/fmicb.2019.00655>.
16. Delgado, E., Fernández-García, A., Pérez-Losada, M., Moreno-Lorenzo, M., Fernández-Miranda, I., Benito, S., et al. (2021). Identification of CRF89_BF, a new member of an HIV-1 circulating BF intersubtype recombinant form family widely spread in South America. *Sci. Rep.* 11, 11442. doi: [10.1038/s41598-021-90023-x](https://doi.org/10.1038/s41598-021-90023-x).
17. Dilernia, D. A., Jones, L. R., Pando, M. A., Rabinovich, R. D., Damilano, G. D., Turk G, et al. (2011). Analysis of HIV type 1 BF recombinant sequences from South America dates the origin of CRF12_BF to a recombination event in the 1970s. *AIDS Res Hum Retroviruses* 27, 569-578. doi: [10.1089/AID.2010.0118](https://doi.org/10.1089/AID.2010.0118).
18. 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](https://doi.org/10.1089/aid.2009.0311).
19. 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; [10.1371/journal.pone.0148549](https://doi.org/10.1371/journal.pone.0148549).
20. Gill, M., Lemey, P., Faria, N., Rambaut, A., Shapiro, B., and Suchard M. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* 30, 713-724. doi: [10.1093/molbev/mss265](https://doi.org/10.1093/molbev/mss265).
21. 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](https://doi.org/10.1097/QAD.0b013e3282f47ad0).
22. 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](https://doi.org/10.1093/sysbio/syq010).

23. Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Kirtley, S., Gouws-Williams, E., et al. (2020). Global and regional epidemiology of HIV-1 recombinants in 1990-2015: a systematic review and global survey. *Lancet HIV* 7, e772-e781. doi: 10.1016/S1473-3099(18)30647-9.
24. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281.
25. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: <https://doi.org/10.1038/nmeth>.
26. 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.
27. Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., et al. (2008). Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J. Infect. Dis.* 197, 707-713. doi: 10.1086/527416.
28. 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.
29. 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.
30. 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. doi: 10.1128/JVI.73.1.152-160.1999.
31. Los Alamos National Laboratory (2021). HIV Sequence Database. <https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>. Retrieved 17 August 2021.

32. Marcelin, A. G., Masquelier, B., Descamps, D., Izopet, J., Charpentier, C., Alloui, C., et al. (2008). Tipranavir-ritonavir genotypic resistance score in protease inhibitor-experienced patients. *Antimicrob. Agents Chemother.* 52, 3237-3243. doi: 10.1128/AAC.00133-08.
33. Nájera, R., Delgado, E., Pérez-Álvarez, L., and Thomson, M. M. (2002). Genetic recombination and its role in the development of the HIV-1 pandemic. *AIDS* 16, S3-S16. doi: 10.1097/00002030-200216004-00002.
34. Paraskevis, D., Nikolopoulos, G. K., Magiorkinis, G., Hodges-Mameletzis, I., and Hatzakis, A. (2016). The application of HIV molecular epidemiology to public health. *Infect. Genet. Evol.* 46, 159-168. doi: 10.1016/j.meegid.2016.06.021.
35. 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.
36. Pessôa, R., Watanabe, J. T., Calabria, P., Felix, A. C., Loureiro, P., Sabino, E. C., et al. (2014a). 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: 10.1093/jac/dkt30510.1371/journal.pone.0112674.
37. Pessôa, R., Carneiro Proietti, A. B., Busch, M. P., and Sanabani, S. S. (2014). Identification of a novel HIV-1 circulating recombinant form (CRF72_BF1) in deep sequencing data from blood donors in Southeastern Brazil. *Genome Announc.* 2, e00386-14. doi: 10.1128/genomeA.00386-14.
38. 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 proviral genomes reveals high genetic diversity among Brazilian blood donors. *PLoS One* 11, e0152499. doi: 10.1371/journal.pone.0152499.
39. 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.

40. 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.
41. Reis, M. N. 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.
42. Reis, M. N. G., Guimarães, 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.
43. Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 31, 298-303. doi: 10.1093/nar/gkg100.
44. Rife, B. D., Mavian, C., Chen, X., Ciccozzi, M., Salemi, M., Min, J., Prosperi, M. C. (2017). Phylodynamic applications in 21st century global infectious disease research. *Glob. Health Res. Policy* 2, 13. doi:10.1186/s41256-017-0034-y.
45. Ristic, N., Zukurov, J., Alkmim, W., Diaz, R. S., Janini, L. M., and Chin, M. P. (2011). Analysis of the origin and evolutionary history of HIV-1 CRF28_BF and CRF29_BF reveals a decreasing prevalence in the AIDS epidemic of Brazil. *PLoS One* 6, e17485. doi: 10.1371/journal.pone.0017485.
46. Ruchansky, D., Casado, C., Russi, J. C., Arbiza, J. R., and López-Galíndez, 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.
47. Sa-Filho, D., Kallas, E. G, Sanabani, S., Sabino, E., Sucupira, M. C., Sanchez-Rosa, A. C., et al. (2007). Characterization of the full-length human immunodeficiency virus-1 genome from recently infected subjects in Brazil. *AIDS Res. Hum. Retroviruses* 23, 1087-1094. doi: 10.1089/aid.2006.0173.
48. Salminen, M. O., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* 11, 1423-1425. doi: 10.1089/aid.1995.11.1423.

49. 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. *Viol. J.* 7, 74. doi:10.1186/1743-422X-7-74.
50. Sanchez, A. M., DeMarco, C. T., Hora, B., Keinonen, S., Chen, Y, Brinkley, C., et al. (2014). Development of a contemporary globally diverse HIV viral panel by the EQAPOL program. *J. Immunol. Methods* 409, 117-130. doi: 10.1016/j.jim.2014.01.004.
51. 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.
52. Seager, I., Travers, S. A., Leeson, M. D., Crampin, A. C., French, N., Glynn, J. R., et al. (2014). Coreceptor usage, diversity, and divergence in drug-naive and drug-exposed individuals from Malawi, infected with HIV-1 subtype C for more than 20 years. *AIDS Res. Hum. Retroviruses* 30, 975-983. doi: 10.1089/aid.2013.0240.
53. Shafer, R. W. (2006). Rationale and Uses of a Public HIV Drug-Resistance Database. *J. Infect. Dis.* 194 Suppl 1, S51-S58. doi: 10.1086/505356.
54. 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.
55. 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.
56. 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.

57. 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.
58. Song, H., Ou, W., Feng, Y., Zhang, J., Li, F., Hu, J., et al. (2019). Disparate impact on CD4 T cell count by two distinct HIV-1 phylogenetic clusters from the same clade. *Proc. Natl. Acad. Sci. U.S.A.* 116, 239-244.
59. 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 1 circulating recombinant form 42 (CRF42_BF) in Luxembourg. *AIDS Res. Hum. Retroviruses* 31, 554-558. doi: 10.1089/AID.2014.0364.
60. 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.
61. Thomson, M. M., Villahermosa, M. L., Vázquez de Parga, E., Cuevas, M. T., Delgado, E., Manjón, N., et al. (2000). Widespread circulation of a B/F intersubtype recombinant form among HIV-1-infected individuals in Buenos Aires, Argentina. *AIDS* 14, 897-899. doi: 10.1097/00002030-200005050-00020.
62. Thomson, M. M., Herrero, I., Villahermosa, M. L., Vázquez de Parga, E., Cuevas, M. T., Carmona, R., 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.
63. Thomson, M. M., Sierra, M., Tanuri, A., May, S., Casado G, Manjón, N., et al. (2004). Analysis of near full-length genome sequences of HIV type 1 BF intersubtype recombinant viruses from Brazil reveals their independent origins and their lack of relationship to CRF12_BF. *AIDS Res. Hum. Retroviruses* 20, 1126-1133. doi: 10.1089/aid.2004.20.1126.
64. Thomson, M. M., and Nájera, R. (2005). Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 7, 210-224.

65. Thomson, M. M., and Fernández-García, A. (2011). Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. *Virology* 415, 30-38. doi: 10.1016/j.virol.2011.03.021.
66. 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.
67. Travers, S. A., Clewley, J. P., Glynn, J. R., Fine, P. E., Crampin, A. C., Sibande, F., et al. (2004). Timing and reconstruction of the most recent common ancestor of the subtype C clade of human immunodeficiency virus type 1. *J. Virol.* 78, 10501-10506. doi: 10.1128/JVI.78.19.10501-10506.2004.
68. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232-W235. doi: 10.1093/nar/gkw256.
69. Vasylyeva, T. I., Zarebski, A., Smyrnov, P., Williams, L. D., Korobchuk, A., Liulchuk, M., et al. (2020) et al. Phylodynamics helps to evaluate the impact of an HIV prevention intervention. *Viruses* 12, 469. doi: 10.3390/v12040469.
70. Venner, C. M., Nankya, I., Kyeyune, F., Demers, K., Kwok, C., Chen, PL, 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.
71. Wensing, A. M., Calvez, V., Ceccherini-Silberstein, F., Charpentier, C., Günthard HF, Paredes, R., et al. (2019). 2019 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 27, 111-121.
72. Zhang, M., Foley, B., Schultz, A. K., Macke, J. P., Bulla, .I, Stanke, M., et al. (2010). The role of recombination in the emergence of a complex and dynamic HIV epidemic. *Retrovirology* 7, 25. doi: 10.1186/1742-4690-7-25.
73. Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391, 594-597. doi: 10.1038/35400.