



# The Origin, Epidemiology, and Phylodynamics of Human Immunodeficiency Virus Type 1 CRF47\_BF

Gracelyn Hill<sup>1</sup>, Marcos Pérez-Losada<sup>1,2,3</sup>, Elena Delgado<sup>4</sup>, Sonia Benito<sup>4</sup>, Vanessa Montero<sup>4</sup>, Horacio Gil<sup>4</sup>, Mónica Sánchez<sup>4</sup>, Javier E. Cañada-García<sup>4</sup>, Elena García-Bodas<sup>4</sup>, Keith A. Crandall<sup>1,2\*</sup>, Michael M. Thomson<sup>4\*</sup> and the Spanish Group for the Study of New HIV Diagnoses

## OPEN ACCESS

### Edited by:

Mariza Gonçalves Morgado,  
Oswaldo Cruz Foundation (FIOCRUZ),  
Brazil

### Reviewed by:

Gonzalo Jose Bello Bentancor,  
Oswaldo Cruz Foundation (FIOCRUZ),  
Brazil  
Leandro Jones,  
Consejo Nacional de Investigaciones  
Científicas y Técnicas (CONICET),  
Argentina

### \*Correspondence:

Keith A. Crandall  
kcrandall@gwu.edu  
Michael M. Thomson  
mthomson@isciii.es

### Specialty section:

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

**Received:** 26 January 2022

**Accepted:** 14 April 2022

**Published:** 16 May 2022

### Citation:

Hill G, Pérez-Losada M,  
Delgado E, Benito S, Montero V,  
Gil H, Sánchez M, Cañada-García JE,  
García-Bodas E, Crandall KA,  
Thomson MM and the Spanish Group  
for the Study of New HIV Diagnoses  
(2022) The Origin, Epidemiology,  
and Phylodynamics of Human  
Immunodeficiency Virus Type 1  
CRF47\_BF.  
Front. Microbiol. 13:863123.  
doi: 10.3389/fmicb.2022.863123

<sup>1</sup> Computational Biology Institute, George Washington University, Washington, DC, United States, <sup>2</sup> Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington, DC, United States, <sup>3</sup> CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Porto, Portugal, <sup>4</sup> HIV Biology and Variability Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain

CRF47\_BF is a circulating recombinant form (CRF) of the human immunodeficiency virus type 1 (HIV-1), the etiological agent of AIDS. CRF47\_BF represents one of 19 CRF<sub>x</sub>\_BFs and has a geographic focus in Spain, where it was first identified in 2010. Since its discovery, CRF47\_BF has expanded considerably in Spain, predominantly through heterosexual contact (~56% of the infections). Little is known, however, about the origin and diversity of this CRF or its epidemiological correlates, as very few samples have been available so far. This study conducts a phylogenetic analysis with representatives of all CRF<sub>x</sub>\_BF sequence types along with HIV-1 M Group subtypes to validate that the CRF47\_BF sequences share a unique evolutionary history. The CRF<sub>x</sub>\_BF sequences cluster into a single, not well supported, clade that includes their dominant parent subtypes (B and F). This clade also includes subtype D and excludes sub-subtype F2. However, the CRF47\_BF sequences all share a most recent common ancestor. Further analysis of this clade couples CRF47\_BF protease-reverse transcriptase sequences and epidemiological data from an additional 87 samples collected throughout Spain, as well as additional CRF47\_BF database sequences from Brazil and Spain to investigate the origin and phylodynamics of CRF47\_BF. The Spanish region with the highest proportion of CRF47\_BF samples in the data set was the Basque Country (43.7%) with Navarre next highest at 19.5%. We include in our analysis epidemiological data on host sex, mode of transmission, time of collection, and geographic region. The phylodynamic analysis indicates that CRF47\_BF originated in Brazil around 1999–2000 and spread to Spain from Brazil in 2002–2003. The virus spread rapidly throughout Spain with an increase in population size from 2011 to 2015 and leveling off more recently. Three strongly supported clusters associated with Spanish regions (Basque Country, Navarre, and Aragon), together comprising 60.8% of the Spanish samples, were identified, one of which was also associated with transmission among men who have sex with men.

The expansion in Spain of CRF47\_BF, together with that of other CRFs and subtype variants of South American origin, previously reported, reflects the increasing relationship between the South American and European HIV-1 epidemics.

**Keywords:** HIV, phylodynamics, circulating recombinant form (CRF), Spain, epidemiology, CRF47\_BF

## INTRODUCTION

High genetic diversity of human immunodeficiency virus type 1 (HIV-1) is a defining feature of the AIDS virus. This diversity gain and loss is a hallmark of the evolution of HIV in the context of drug resistance and changing environments (Pennings et al., 2014). A contributing factor in the evolution of HIV is the process of recombination (Rambaut et al., 2004; Vuilleumier and Bonhoeffer, 2015). Genetic recombination is known to impact HIV allelic diversity and subsequent population dynamics at a rate equivalent to the high mutation rate of HIV (Shriner et al., 2004). Genetic diversity within HIV subtypes can be up to 17% sequence divergence across the genome with 17–35% divergence between subtypes (Castro-Nallar et al., 2012a). Yet recombination can even occur between subtypes as HIV variants spread around the globe, leading to circulating recombinant forms or CRFs, as well as unique recombinant forms (URFs) (Castro-Nallar et al., 2012b). There are currently 118 known HIV-1 CRFs according to the Los Alamos HIV Sequence Database (Los Alamos National Laboratory, 2021) involving recombination events between nearly all known subtypes and even between other CRFs [e.g., CRF15\_01B is a recombinant form between CRF01 and subtype B (Tovanabutra et al., 2003)]. The CRFs often have their own unique population dynamics and molecular epidemiology compared to their parental strains and often lead to novel infection dynamics and spread. One such CRF is CRF47\_BF, discovered in Spain and described in 2010 (Fernández-García et al., 2010) as an intersubtype recombinant form between HIV-1 subtypes B and F. Of the CRFs, among the most abundant are those between B and F subtypes, with 19 CRF\_BFs (note that in the Los Alamos HIV Database these are sometimes designated “BF” and sometimes “BF1,” even for the same CRF). Of the CRF\_BFs, all but two are known from South America (mainly Brazil, but Argentina, Uruguay, Paraguay, Chile, Peru, and Bolivia as well) with a few found both in South America and Europe (CRF66, 75, and 89). Only two CRF\_BF have been reported to be found circulating exclusively in Europe, CRF42\_BF in Luxembourg (Struck et al., 2015) and CRF47\_BF in Spain (Fernández-García et al., 2010). Since its description, CRF47\_BF has expanded considerably in Spain, predominantly *via* heterosexual contact, and is now known from Brazil as well, as attested by a CRF47\_BF virus collected in this country whose sequence is deposited in the Los Alamos database (Los Alamos National Laboratory, 2021).

The goal of this study is to estimate the temporal and geographic origin of CRF47\_BF and the dynamics of diffusion and growth throughout its evolutionary history. Toward this goal, we combine new CRF47\_BF sequence data from our lab from strains isolated in Spain with data from other BF strains in

the Los Alamos database to examine the origin and evolutionary dynamics of CRF47\_BF and their epidemiological correlates.

## MATERIALS AND METHODS

### Sample and Data Collection

Plasma and whole blood samples were collected from HIV-1-infected patients at public hospitals across eight regions in Spain for a molecular epidemiological study of all new HIV-1 diagnoses seen at the participating centers and for antiretroviral drug resistance testing. Epidemiological data from the CRF47\_BF patients were collected to link to the HIV sequence data. The epidemiological data included patient gender, the transmission route, the patient's year of HIV diagnosis and date of sample collection, the region from which the sample was collected, the country of origin of the individual, and whether the patient was on antiretroviral (ARV) therapy.

The study was approved by the Committee of Research Ethics of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain (report numbers CEI PI 38\_2016-v3 and CEI PI 31\_2019-v5). The study 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.

### Sequence Analyses

(RT-)PCR was used to amplify the protease-reverse transcriptase (PR-RT) gene region from plasma-extracted RNA or whole blood-extracted DNA using previously described primers (Delgado et al., 2015; **Supplementary Figure 1**). PCR products were sequenced using the Sanger method with an automated capillary sequencer. These data were combined with PR-RT sequences classified as CRF47\_BF at the Los Alamos HIV Sequence Database and reference sequences for all subtypes and all CRF<sub>x</sub>\_BFs for this same gene region from the Los Alamos HIV Database. Finally, we conducted a BLAST (Altschul et al., 1990) search against GenBank with the 5'-most 950 nt of PR-RT of all CRF47\_BF viruses and included all sequences within 95% similarity. BLAST searches and further analyses (see below) yielded only two additional CRF47\_BF sequences not identified as such at the Los Alamos database (with GenBank accessions JF929086, from Spain, and JQ238096, from Brazil).

We conducted two analyses with these data. (1) We included all data to validate the quality of the data and place the CRF47\_BF within a broader phylogenetic context. Our initial phylogenetic analysis included subtypes from the HIV-1 M group (subtypes A1, A2, B, C, D, F1, F2, G, H, J, K, and L), as well as the CRF\_BF recombinants (TotalCRF\_BF.fasta, see **Supplementary Material**). Our final alignment (1,200 bp) included 14 sequences

representing all the major subtypes within HIV-1 group M, 5 subtype B sequences, 11 subtype F (F1, F2) sequences, and 34 representatives of all known and distinct CRF\_BFs. This alignment also included our more focused (2) CRF47\_BF dataset (CRF47\_BF.fasta see **Supplementary Material**). By including additional subtypes (including lab strains), we can both verify the monophyly of our target group of CRF47\_BF sequences and validate that there are not contaminants or strange recombinants within this group as would be indicated by novel phylogenetic placement. For this second dataset, we included all 99 sequences from CRF47\_BF, including 87 obtained by us [7 from a previous study (Fernández-García et al., 2010) and 80 newly derived] from the patients summarized in **Table 1**, and 12 from databases (10 from Spain and 2 from Brazil). We then conducted a focused analysis on the targeted CRF47\_BF strains (1,377 aligned bp).

In both analyses, we aligned sequence data using MAFFT (Katoh and Standley, 2013) with the FFT-NS-2 progressive alignment approach since these sequences are relatively similar. Prior to subsequent phylogenetic and phylodynamic analyses, we checked that all sequences showed mosaic structures coincident with CRF47\_BF, through two procedures: (1) bootscan analyses

and (2) separate phylogenetic trees of B and F1 segments previously defined for CRF47\_BF (Fernández-García et al., 2010), including B and F1 subtype references, to ensure that subtype assignment of each segment was identical for all sequences. Phylogenetic analyses were conducted using maximum-likelihood (Felsenstein, 1981; Posada and Crandall, 2021) as implemented by RAxML (Kozlov et al., 2019) *via* the CIPRES web service (Miller et al., 2012). The phylogenetic analyses utilized the best-fit model of evolution (Posada and Crandall, 1998) as determined by ModelTest-NG (Darriba et al., 2020). Phylogenetic analyses were also done using a Bayesian approach as implemented by MrBayes 3.2 (Ronquist et al., 2012) with integrated model selection, 10 million MCMC generations, and codon partitioning. Confidence in the resulting phylogenetic estimates was assessed using the bootstrap approach (Felsenstein, 1985) for the maximum-likelihood analyses with 1,000 pseudoreplicates and with posterior probabilities (pP) in the Bayesian framework. Phylogenetic trees were visualized with iTOL (Letunic and Bork, 2019), as well as mapping of epidemiological characters along the phylogeny.

We applied BEAST2 (Bouckaert et al., 2014) to the CRF47\_BF dataset to estimate a chronogram and the phylodynamic history of CRF47\_BF. First, we validated the existence of temporal signal in the dataset with TempEst v1.5.3 (Rambaut et al., 2016), which determines the correlation of genetic divergence among sequences (measured as root-to-tip distance) with time. For the BEAST2 analysis we ran 10 million generations, two codon partitions (1st + 2nd, and 3rd positions), used an uncorrected log-normal relaxed molecular clock (initial ucl.d.mean = 1.0 and initial ucl.d.stdev = 0.333), estimated base frequencies and the HKY + G evolution model. The input file was created using BEAUti. Past population dynamics was estimated *via* Skygrid analysis (Hill and Baele, 2019) using a coalescent Bayesian Skygrid tree prior. We used Tracer (Rambaut et al., 2018) to verify convergence and to visualize the Skygrid plot. We compare the inferred effective population size of the CRF47\_BF population in Spain to the proportion of CRF47\_BF diagnoses over time across the same study regions and time period.

Finally, known drug resistant mutations were identified in the focused CRF47\_BF data using the Stanford HIV Drug Resistance Database's HIVdb v9.0 program (Tang et al., 2012).

## Statistical Analyses

Correlations between cluster membership and epidemiological data were analyzed with Fisher's exact test.

## RESULTS

### Epidemiology and Sequences

We collected samples and epidemiological data from 87 patients throughout eight different regions of Spain (Basque Country, Navarre, Galicia, Aragon, Comunitat Valenciana, Madrid, Castilla-La Mancha, and Castilla y León) (**Figure 1**). Collections were made from 2007 to 2021. Males accounted for 78% of the individuals with CRF47\_BF in our study and 56% of individuals reported transmission *via* heterosexual contact (61%

**TABLE 1** | Summary data from patients with CRF47\_BF variant from Spain.

|   | Total N = 87 | Percent | New sequences |
|---|--------------|---------|---------------|
| <b>Gender</b>                               |              |         |               |
| Male  | 68           | 78.2    |               |
| Female                                      | 18           | 20.7    |               |
| Transgender                                 | 1            | 1.2     |               |
| <b>Region</b>                               |              |         |               |
| Basque Country                              | 38           | 43.7    | 35 (38 total) |
| Navarre                                     | 17           | 19.5    | 17            |
| Galicia                                     | 14           | 16.1    | 10 (14 total) |
| Aragon                                      | 9            | 10.3    | 9             |
| Comunitat Valenciana                        | 6            | 6.9     | 6 (13 total)  |
| Castilla y León                             | 1            | 1.2     | 1             |
| Castilla-La Mancha                          | 1            | 1.2     | 1             |
| Madrid                                      | 1            | 1.2     | 1             |
| <b>Transmission route</b>                   |              |         |               |
| Heterosexual                                | 49           | 56.3    |               |
| MSM   | 19           | 21.8    |               |
| Sexual transmission (unspecified sexuality) | 12           | 13.8    |               |
| Other/no data                               | 7            | 8.1     |               |
| <b>ARV therapy</b>                          |              |         |               |
| No  | 75           | 86.2    |               |
| Yes   | 7            | 8.1     |               |
| No data                                     | 5            | 5.7     |               |
| <b>Country of origin</b>                    |              |         |               |
| Spain                                       | 68           | 78.2    |               |
| Brazil                                      | 5            | 5.8     |               |
| Colombia                                    | 5            | 5.8     |               |
| Morocco                                     | 3            | 3.5     |               |
| Nicaragua                                   | 2            | 2.3     |               |
| Other                                       | 4            | 4.6     |               |

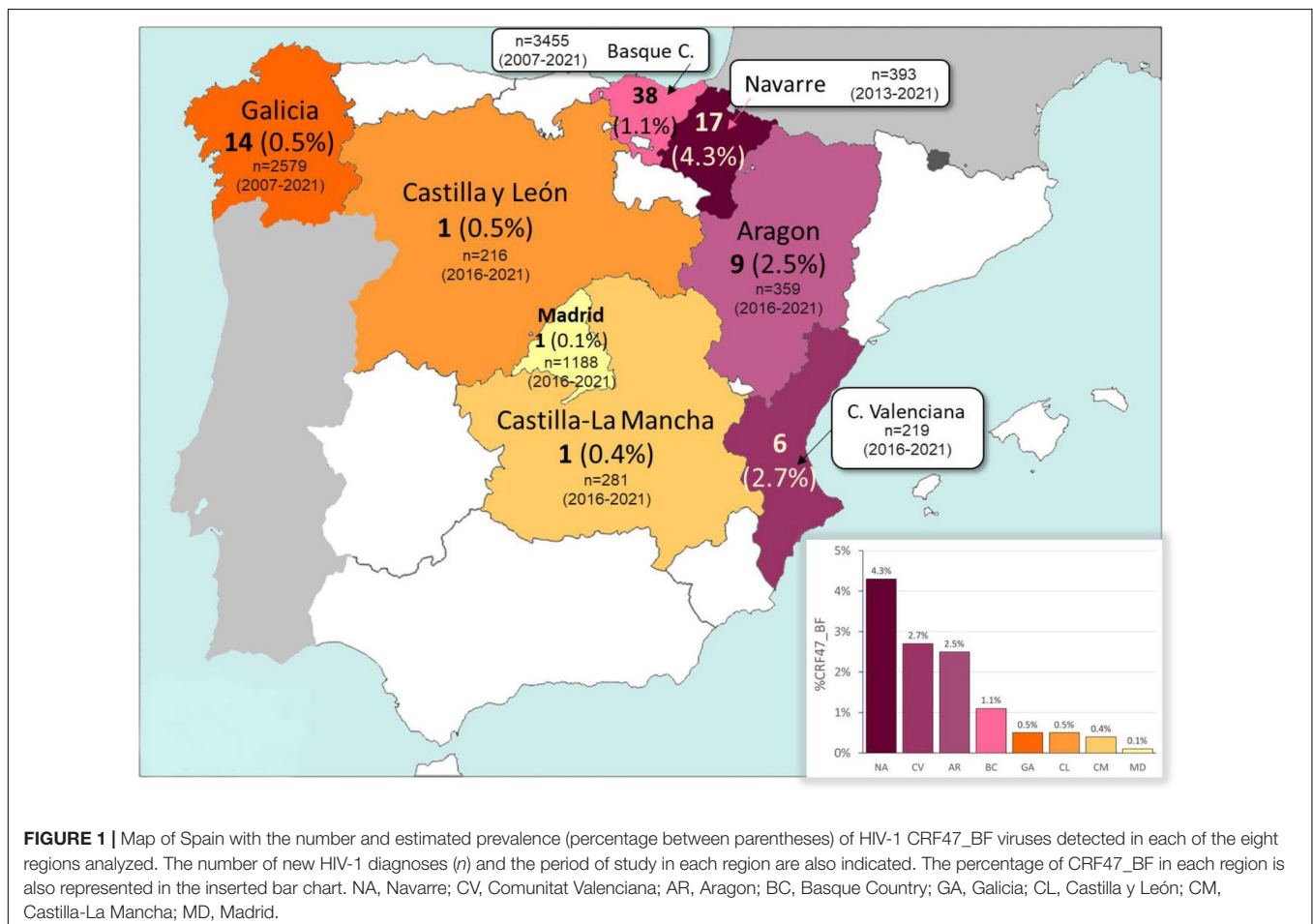
considering only individuals with available data on transmission route) (Table 1). The Spanish region with the highest proportion of the CRF47\_BF variant in our data set was the Basque Country with 44% of the cases, while Navarre was the next highest (18% of the cases) (see Table 1 for number of new CRF47\_BF sequences and total CRF47\_BF sequences per region, and Figure 1 for the total number of analyzed HIV-1 sequences and prevalence of CRF47\_BF among new HIV-1 diagnoses in each region in the sampling periods). Most samples were collected shortly after HIV diagnosis. Patients received ARV therapy after sample collection.

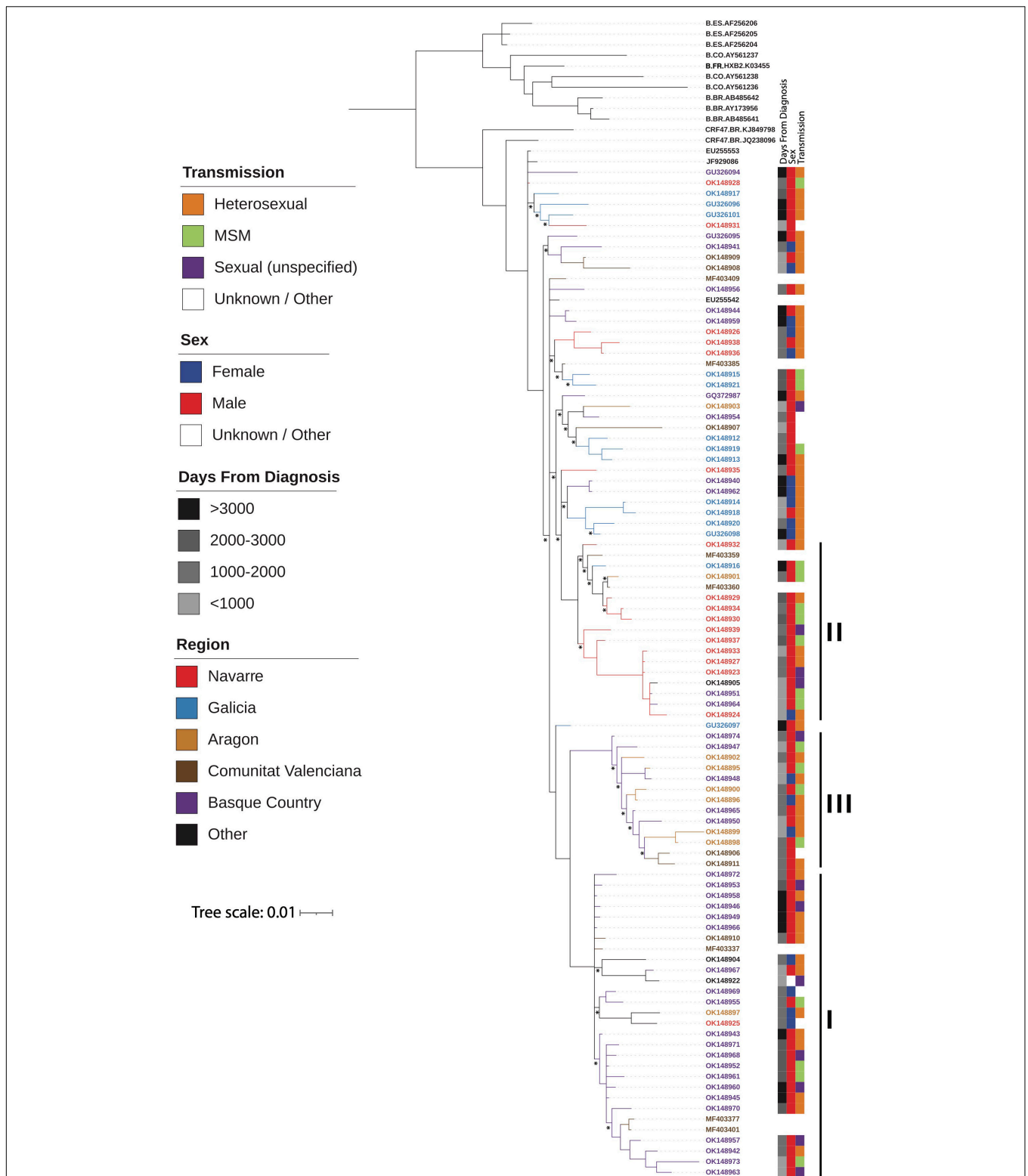
## Phylogenetics

The first phylogenetic analysis was a maximum likelihood phylogenetic estimate of the relationships amongst the CRF<sub>x</sub>BFs, including HIV-1 M subtypes as outgroup taxa and subtypes B, F, and CRF<sub>x</sub>BFs as ingroup taxa. Our RAXML tree depicted a monophyletic cluster of the subtype B, F, and CRF<sub>x</sub>BFs relative to the other HIV-1 subtypes (Supplementary Figure 2), but including also Subtype D. The backbone structure of the CRF phylogenetic relationships was weakly supported (<70% bootstrap support – indicated by dashed lines), which is not particularly surprising given the potential difficulty in representing evolutionary histories of recombinant HIV-1 forms as bifurcating trees (Posada and Crandall, 2001, 2002). Many of

the CRF<sub>x</sub>BF forms cluster in strongly supported monophyletic groups themselves (e.g., CRF40\_BF, CRF72\_BF, CRF75\_BF, CRF90\_BF, CRF89\_BF, etc.), including our target group of CRF47\_BF sequences. Many of the other CRFs form weakly supported monophyletic groups (e.g., CRF70\_BF, CRF46\_BF, CRF38\_BF, etc.) and a few form non-monophyletic groupings (e.g., CRF66\_BF and CRF71\_BF). The subtype B sequences cluster together within the CRF<sub>x</sub>BF clade with both a cluster of subtype D and the CRF28\_BF sequence nested within this subtype B cluster. Nevertheless, the target group for this study, the CRF47\_BF sequences, clearly form a monophyletic group, suggesting independent evolution, and are a sister group to the CRF44\_BF clade.

The Bayesian estimated phylogeny for the CRF47\_BF sequences shows a monophyletic grouping of the sequences from Spain (Figure 2) with the two sequences from Brazil (KJ849798 and JQ238096) branching basally. Within the Spanish cluster, there are three strongly supported clusters, comprising 29 (cluster I), 17 (cluster II), and 13 (cluster III) viruses, respectively, which are associated with the Basque Country ( $p = 0.0002$ ), Navarre ( $p = 0.0001$ ), and Aragon ( $p = 0.0002$ ), respectively. This is indicative of a single introduction of CRF47\_BF into Spain with subsequent spread throughout the country and point introductions with subsequent expansion in different regions.





**FIGURE 2 |** Majority-rule consensus Bayesian (MrBayes) phylogenetic estimate of CRF47\_BF sequences from Spain (colored by region) and other CRF47\_BF sequences from GenBank, as well as a few additional subtype B sequences from Spain, Brazil, and Colombia plus HXB2, all to serve as an outgroup to the CRF47 sequences. Only clade posterior probabilities <math><0.95</math> are indicated by an \*; all other clades showed posterior probabilities  $\geq 0.95$ . Epidemiological data are mapped to the right of the phylogeny, including days from diagnosis, sex, transmission, and geographic region. Branch lengths are drawn proportional to the amount of sequence divergence. Clusters corresponding to the Spanish regions of Basque Country (cluster I), Navarre (cluster II), and Aragon (cluster III) are indicated.

The mixing of patient gender throughout the resulting phylogeny supports the epidemiological data suggesting predominantly heterosexual transmission among patients. We also found that cluster II, associated with Navarre, was associated with men who have sex with men (MSM) ( $p = 0.0388$ ). In this cluster, 14 of 15 individuals with known gender are men.

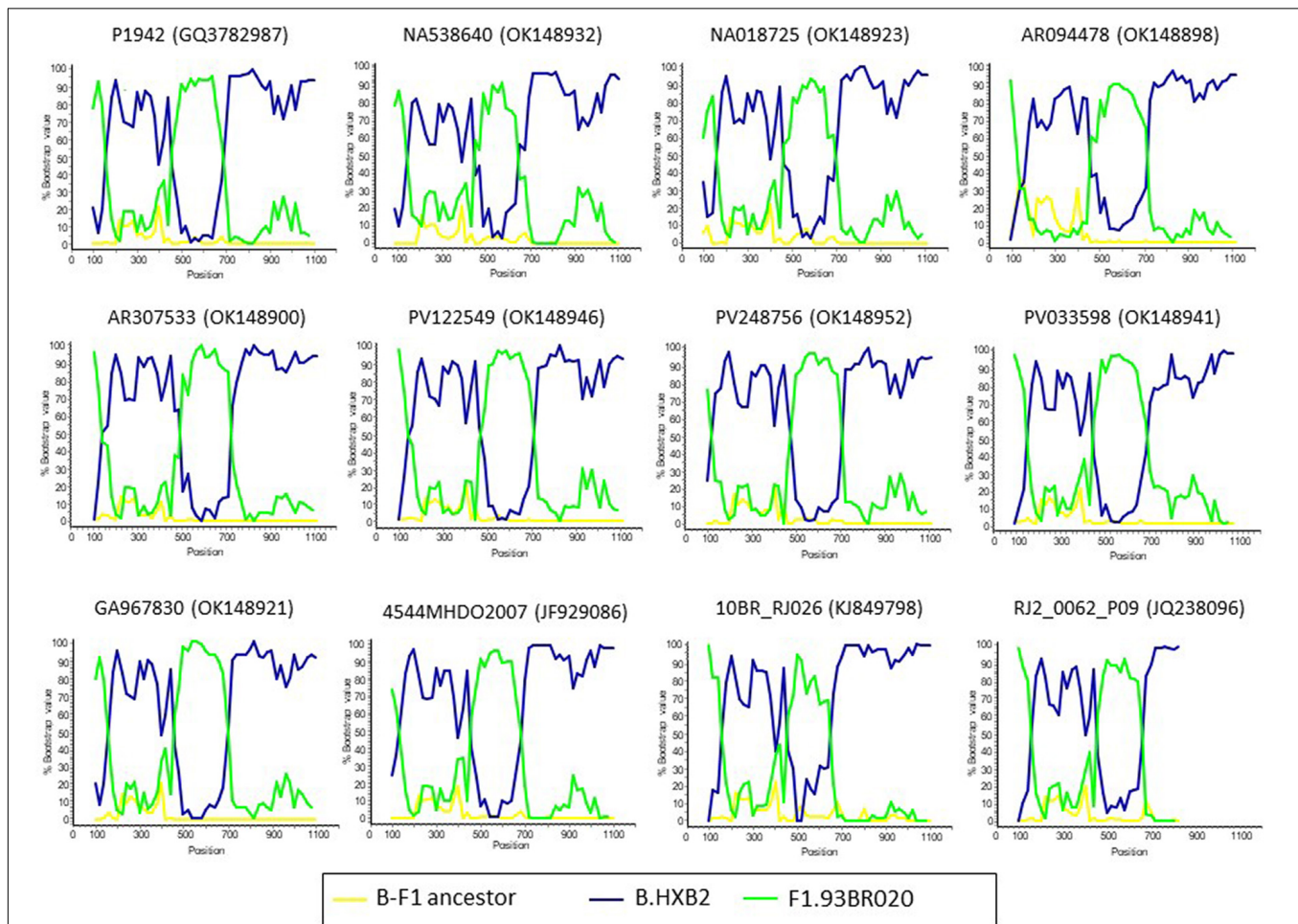
The bootscan analysis for recombination and separate trees of B and F segments suggest that all the target sequences within the CRF47\_BF analyses presented here share the same recombination pattern (as outlined at the Los Alamos HIV Database) (Figure 3 and Supplementary Figure 3). Thus, while recombination can significantly impact phylogenetic interpretations (certainly, for the overall tree presented in Supplementary Figure 2), it does not seem to be differentially impacting analyses of our targeted CRF47\_BF sequences.

Based on the sample dates, we grouped these in four temporal categories of recency (days between diagnosis date and current date) ( $> 3,000$ ,  $2,000-3,000$ ,  $1,000-2,000$ , and  $< 1,000$  days from current) for ease of visualization of time over the phylogeny

to test for temporal clustering. Thus, the greater the value the closer to the most recent common ancestor, i.e., origin of CRF47\_BF. Note that these correspond well to the branch lengths observed leading to samples with  $< 1,000$  days from diagnosis having longer branches from the root to the terminal samples and  $> 3,000$  having shorter and more basal branches in the phylogram. No temporal clustering was observed as these different time categories were distributed throughout the CRF47\_BF phylogeny (Figure 2).

## Analysis of Drug Resistance Mutations

To identify drug resistance mutations in the CRF47\_BF viruses, we analyzed the sequences with the Stanford HIV Drug Resistance Database's HIVdb program (Tang et al., 2012). We found ARV drug resistance mutations in five patients: M184V or M184I mutations of resistance to nucleoside reverse transcriptase inhibitors (NRTI) in three samples; K103N mutation of resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI) + K65N mutation of resistance to NRTIs in one sample;



**FIGURE 3 |** Bootscan analyses of PR-RT sequences of CRF47\_BF viruses. Simplot v3.5 (Lole et al., 1999) was used for the analyses. Twelve representative profiles are displayed. Names of viruses, with GenBank accessions, are shown above each bootscan plot. P1942 was included as CRF47\_BF reference. A reconstructed B-F1 ancestral sequence was used as outgroup. The horizontal axis represents the position from nucleotide 1 of protease and the vertical axis represents bootstrap values supporting clustering with references.

and E138A mutation associated with low level resistance to the NNRTI rilpivirine in one patient. Only one of these patients, with M184I mutation, was ARV drug-experienced.

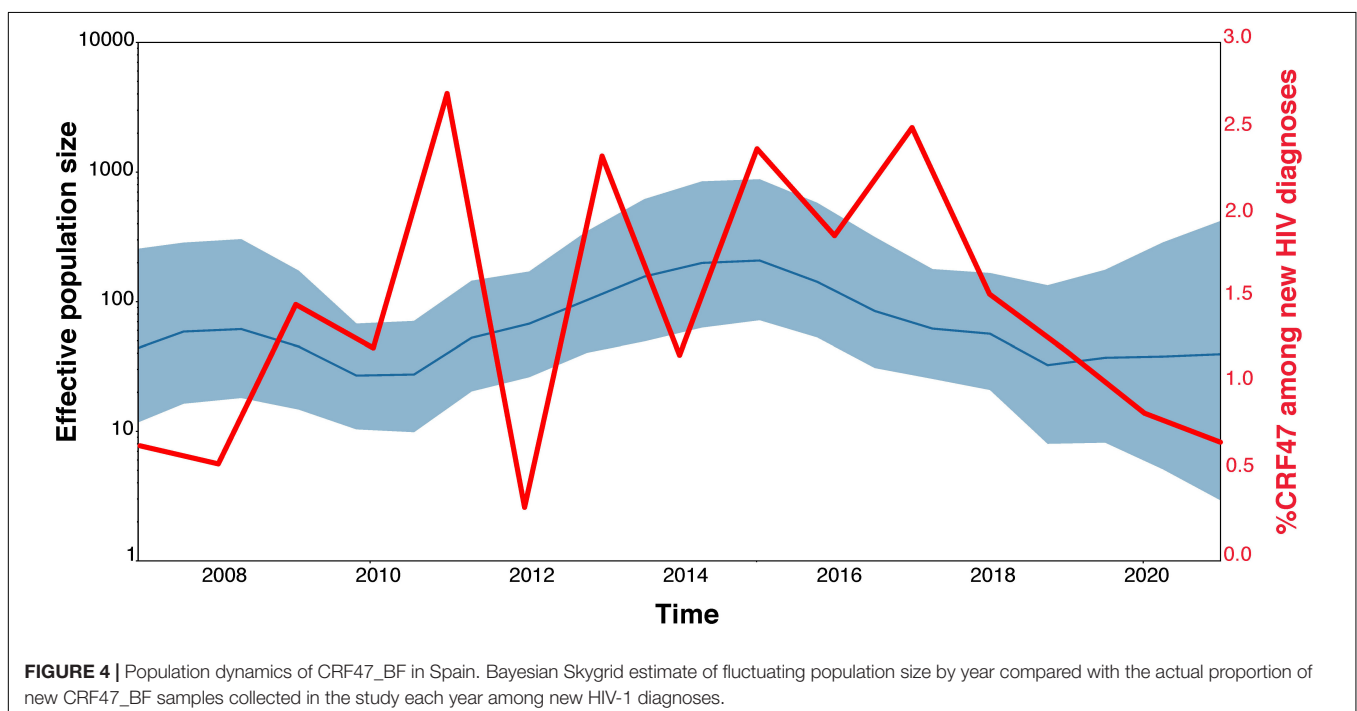
## Phylodynamics

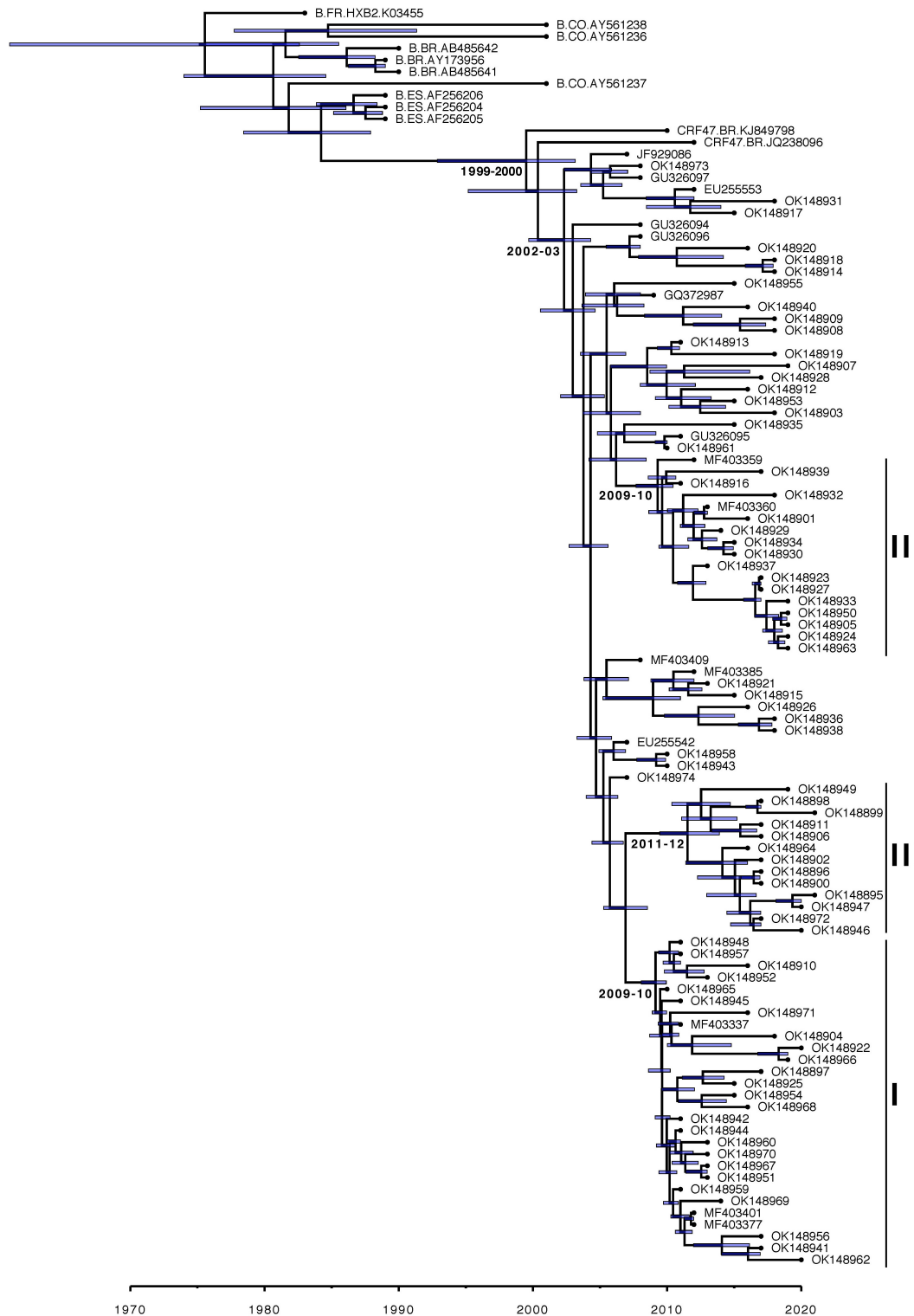
Our TempEst analysis determined that there was an adequate temporal signal in the dataset ( $R^2 = 0.5051$ ). With time-stamped sequence data, we performed a Bayesian Skygrid coalescent analysis to estimate historical population dynamics (Hill and Baele, 2019) of the CRF47\_BF variants throughout Spain. Time labels (tipdates) were determined by the date of sample collection (ranging from 2007 to 2021). Our analysis supports a fairly dynamic population history of the CRF47\_BF in Spain over the last 15 years with an initial increase in population size, a subsequent increase from 2011 to 2015, with a leveling off more recently, but seemingly increasing variance (Figure 4). This fluctuation in effective population size of CRF47\_BF in Spain is not as dynamic as the % CRF47\_BF infections among new HIV diagnoses, that fluctuate considerably over this same time period (Figure 4), but with similar overall trends. The average effective population size was estimated to be 155 with a mean substitution rate of  $1.8128 \times 10^{-3}$  [95% highest posterior density (HPD) interval ( $1.3956 \times 10^{-3}$ ,  $2.2548 \times 10^{-3}$ )]. Using BEAST, we estimated a chronogram to determine the time of origin for both the CRF47\_BF clade as well as the timing of the introduction of CRF47\_BF viruses to Spain (Figure 5). We estimated the origin of the CRF47\_BF clade in Brazil (pP = 1.0), dated to 1999–2000 (95% HPD interval between 1994 and 2003) and timed the introduction of CRF47\_BF to Spain (pP = 0.99) to be 2002–2003 (95% HPD interval between 2000 and 2004) (Figure 5). Similarly, viral strains seem to have entered once and spread through the Spanish regions

of Basque Country (cluster I) (pP = 1.0), Navarre (cluster II) (pP = 1.0), and Aragon (cluster III) (pP = 1.0) between 2009 and 2012 (Figure 5). These analyses, hence, suggest that CRF47\_BF was probably circulating in Spain for about 8 years before it was identified through DNA sequencing, but clearly at a relatively low frequency. Given the sampling of CRF47 sequences, it appears that the introduction of this recombinant form to Spain was *via* Brazil, supported by very high posterior probabilities (pP = 1.00).

## DISCUSSION

The HIV-1 CRF47\_BF was first reported in 2010, detected in nine samples collected in Spain in 2007–2009. Samples have subsequently been collected as this novel variant has spread throughout the country. Our phylogenetic analysis shows that isolates of CRF47\_BF form a strongly supported monophyletic group (share a most recent common ancestor) relative to other CRFx\_BF sequences, distinct from other CRFx\_BF sequences, subtype B, sub-subtype F2 and other Group M subtypes. A focused phylogenetic analysis of the CRF47\_BF sequences show a clear single origin in Brazil around 1999–2000 with a subsequent transmission and rapid spread throughout Spain beginning around 2002–2003. Three strongly supported clusters, comprising a majority of viruses and associated with the regions of Basque Country, Navarre, and Aragon, were identified; this suggests that after a single introduction in Spain, CRF47\_BF has spread mainly through localized point introductions and subsequent spread in different geographical areas. CRF47\_BF is predominant in males (78%) with a predominantly heterosexual transmission (56% of the total, 61%





**FIGURE 5 |** Maximum clade credibility Bayesian (BEAST) chronogram estimate of CRF47\_BF sequences from Spain and other CRF47\_BF sequences from GenBank, as well as a few additional subtype B sequences from Spain, Brazil, and Colombia plus HXB2, all to serve as an outgroup to the CRF47\_BF sequences. Clusters associated with the Spanish regions of Basque Country (cluster I), Navarre (cluster II), and Aragon (cluster III) are indicated. Estimated years of emergence of CRF47\_BF in Brazil, of its introduction in Spain, and of emergence of the Spanish clusters are indicated besides the corresponding nodes. A total of 95% highest posterior density (HPD) intervals (blue bars) are shown for all time estimates.



of those with data on transmission mode). The phylodynamic analysis and percent of CRF47 among new HIV diagnoses both support a fluctuating population size of CRF47\_BF over the last 15 years with periods of expansion and contraction, suggesting that continued monitoring of this novel variant will be important to track its spread.

It is interesting to note that one cluster of 17 individuals, associated with Navarre, where 14 of 15 individuals with available data were male, was significantly associated ( $p = 0.0388$ ) with transmission among MSM. Although three men were reported to be heterosexual, considering the great male preponderance in the cluster, it is probable that they are non-disclosed MSM (Hué et al., 2014; Ragonnet-Cronin et al., 2018). The identification of an MSM-associated cluster within the CRF47\_BF clade may be indicative of the diffusion of CRF47\_BF from a heterosexual-driven network to a MSM-driven network. A similar phenomenon has been observed for the two other CRFs of South American origin identified by us in Spain: CRF66\_BF (Bacqué et al., 2021) and CRF89\_BF (Delgado et al., 2021). Such phenomenon may reflect the migration of these CRFs from countries where heterosexual transmission is predominant to Spain, where most currently expanding HIV-1 clusters are associated with MSM (Patiño-Galindo et al., 2017; Gil et al., 2022). It should be pointed out, however, that outside of the Navarre cluster, the male:female ratio was 3.2:1, which contrasts to the 2.4:1 ratio of self-declared heterosexual men to MSM (decreasing to 1.5:1 if all men with non-specified sexual transmission were MSM). This discrepancy could also be explained by the presence of non-disclosed MSM among self-declared heterosexual men outside of the Navarre cluster, and indicates that epidemiological data on transmission route based on self-reported sexual behaviors should be interpreted with caution.

The recent expansion in Spain of CRF47\_BF, whose Brazilian origin is first reported here, is one more example of the increasing relationship of the South American and European HIV-1 epidemics, also reflected in the propagation in Europe of other CRFs (12\_BF, 17\_BF, 60\_BC, 66\_BF, and 89\_BF) (Simonetti et al., 2014; Fabeni et al., 2015, 2020; Bacqué et al., 2021; Delgado et al., 2021) and variants of subtypes F1 and C (Tovanabutra et al., 2003; de Oliveira et al., 2010; Thomson et al., 2012; Lai et al., 2014; Carvalho et al., 2015; Delgado et al., 2015; Vinken et al., 2019) of South American ancestry, which probably derives from increasing migratory flows from South America to Europe.

The repeated introduction and expansion in Spain of multiple CRFs and non-B subtypes (Delgado et al., 2015, 2019; Patiño-Galindo et al., 2017; González-Domenech et al., 2018; Kostaki et al., 2019) justifies the establishment of a HIV-1 molecular epidemiological surveillance system, aimed at promptly detecting the propagation of such variants, as well as rapidly expanding clusters, that could provide information in real-time on changes in the genetic composition and the dynamics of the HIV-1 epidemic to guide the implementation of preventive public health interventions (Paraskevis et al., 2016; German et al., 2017; Oster et al., 2018). Nevertheless, phylogenetic analyses of HIV sequence data should be taken with caution as recombination can impact phylogenetic inference (Schierup and Hein, 2000; Posada and

Crandall, 2002) suggesting network approaches might be better suited for representation of such data (Clement et al., 2000). Nevertheless, we focus on a specific set of CRF47\_BF sequences with a shared mosaic structure and therefore our results should be robust to the impacts of recombination (see **Figure 3**).

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, OK148895-OK148974.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Committee of Research Ethics of Instituto de Salud Carlos III, Majadahonda, Madrid, Spain (report numbers CEI PI 38\_2016-v3 and CEI PI 31\_2019-v5). Written informed consent for participation was not required for this study, as it used samples and data collected as part of routine clinical practice and patients' data were anonymized without retaining data allowing individual identification.

## MEMBERS OF THE SPANISH GROUP FOR THE STUDY OF NEW HIV DIAGNOSES

Hospital Universitario de Basurto: José Luis Díaz de Tuesta del Arco, Silvia Hernández, Sofía Ibarra-Ugarte, Josefa Muñoz, María Carmen Nieto-Toboso, and Miren Zuriñe Zubero-Sulibarria. Hospital Universitario de Cruces, Bilbao: Elena Bereciartua-Bastarrica, Luis Elorduy, Ane Josune Goikoetxea-Agirre, and Leyre López-Soria. Hospital Universitario de Galdakao: María José López de Goikoetxea. Hospital Universitario Donostia, San Sebastián: Maitane Aranzamendi, Gustavo Cilla, José Antonio Iribarren, and Yolanda Salicio. Hospital Universitario Araba, Vitoria: Carmen Gómez, and José Joaquín Portu. Hospital Universitario de Navarra, Pamplona: Aitziber Aguinaga, Carmen Ezpeleta, Carmen Martín-Salas, and María Gracia Ruiz-Alda. Hospital Reina Sofía, Tudela: José Javier García-Irure. Hospital Universitario Sant Joan d'Alacant: Fernando Buñuel and Francisco Jover-Díaz. Complejo Hospitalario Universitario de Vigo: Jorge Julio Cabrera, Antonio Ocampo, and Celia Miralles. Complejo Hospitalario Universitario de Pontevedra: Julio Diz-Aren and Matilde Trigo. Complejo Hospitalario Lucus Augusti, Lugo: María José Gude, Ramón Rabuñal, and Eva María Romay. Complejo Hospitalario Universitario de Ourense: Ricardo Fernández-Rodríguez and Juan García-Costa. Hospital Universitario Miguel Servet, Zaragoza: Piedad Arazo and Ana María Martínez-Sapiña. Centro Sanitario Sandoval, Madrid: Jorge del Romero. Hospital Universitario Río Hortega, Valladolid: Belén Lorenzo-Vidal. Hospital Universitario de Toledo: César Gómez.

## AUTHOR CONTRIBUTIONS

MT, ED, and MP-L conceived of the project. ED collected sequence data from the samples. GH, KC, MP-L, MT, and ED conducted the data analyses. HG performed the data curation. SB, VM, MS, JC-G, and EG-B performed the experimental work. The members of the Spanish Group for the Study of New HIV Diagnoses collected the samples and clinical and epidemiological data for the study. KC, GH, and MP-L wrote the original draft of the manuscript. MT, ED, and HG edited the manuscript. All authors read and approved the manuscript.

## FUNDING

The study was supported by Acción Estratégica en Salud Intramural (AESI) program of Instituto de Salud Carlos III, projects “Estudio sobre Vigilancia Epidemiológica Molecular de la Infección por 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; Red de Investigación en SIDA (RIS), Instituto de Salud Carlos

III, Plan Nacional I+D+I, project RD16ISCIII/0002/0004; and scientific agreements with the Governments of Galicia (MVI 1004/16) and Basque Country (MVI 1001/16).

## ACKNOWLEDGMENTS

We thank José Antonio Taboada, from Consellería de Sanidade, Government of Galicia, and Daniel Zulaika, from Unidad de Coordinación del Plan de Prevención y Control del SIDA, Osakidetza-Servicio Vasco de Salud, Government of the Basque Country, 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.2022.863123/full#supplementary-material>

## REFERENCES

- 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.
- 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
- 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
- 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.
- Castro-Nallar, E., Crandall, K. A., and Pérez-Losada, M. (2012a). Genetic diversity and molecular epidemiology of HIV transmission. *Future Virol.* 7, 239–252.
- Castro-Nallar, E., Pérez-Losada, M., Burton, G. F., and Crandall, K. A. (2012b). The evolution of HIV: inferences using phylogenetics. *Mol. Phylogenet. Evol.* 62, 777–792.
- Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659.
- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294.
- de Oliveira, T., Pillay, D., Gifford, R. J., and UK Collaborative Group on HIV Drug Resistance (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
- 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: 10.3389/fmicb.2019.00655
- 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., 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.
- 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
- Fabeni, L., Santoro, M. M., Lorenzini, P., Rusconi, S., Gianotti, N., Costantini, A., et al. (2020). Evaluation of HIV transmission clusters among natives and foreigners living in Italy. *Viruses* 12:791.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- 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.
- 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
- Gil, H., Delgado, E., Benito, S., Georgalis, L., Montero, V., Sánchez, M. et al., (2022). Transmission clusters, predominantly associated with men who have sex with men, play a main role in the propagation of HIV-1 in northern Spain (2013–2018). *Front. Microbiol.* 13:782609. doi: 10.3389/fmicb.2022.782609
- González-Domenech, C. M., Viciano, 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-naïve patients in southern Spain. *PLoS One* 13:e0190544. doi: 10.1371/journal.pone.0190544
- Hill, V., and Baele, G. (2019). Bayesian estimation of past population dynamics in BEAST 1.10 using the Skygrid coalescent model. *Mol. Biol. Evol.* 36, 2620–2628. doi: 10.1093/molbev/msz172
- 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
- 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.
- Kostaki, E. G., Flampouris, A., Karamitros, T., Chueca, N., Alvarez, M., Casas, P., et al. (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
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455.
- 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.
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadhari, 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
- Los Alamos National Laboratory [LANL] *HIV Databases*. Available online at: <https://www.hiv.lanl.gov/content/index> (accessed March 20, 2021).
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2012). “The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources,” in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond XSEDE '12*, (New York, NY: Association for Computing Machinery), 1–8.
- Oster, A. M., France, A. M., and Mermin, J. (2018). Molecular epidemiology and the transformation of HIV prevention. *JAMA* 319, 1657–1658. doi: 10.1001/jama.2018.1513
- 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
- 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
- Pennings, P. S., Kryazhimskiy, S., and Wakeley, J. (2014). Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet.* 10:e1004000. doi: 10.1371/journal.pgen.1004000
- Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Posada, D., and Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16, 37–45.
- Posada, D., and Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J. Mol. Evol.* 54, 396–402.
- Posada, D., and Crandall, K. A. (2021). Felsenstein phylogenetic likelihood. *J. Mol. Evol.* 89, 134–145.
- 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
- 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.
- 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:vev007.
- 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.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
- Schierup, M. H., and Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879–891.
- 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.
- 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.
- 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.
- Tang, M. W., Liu, T. F., and Shafer, R. W. (2012). The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology* 55, 98–101.
- 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, e49–51.
- Tovanabutra, S., Watanaveeradej, V., Viputtikul, K., De Souza, M., Razak, M. H., Suriyanon, V., et al. (2003). A new circulating recombinant form, CRF15\_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand. *AIDS Res. Hum. Retroviruses* 19, 561–567.
- Vinken, L., Fransen, K., Cuyppers, L., Alexiev, I., Balotta, C., Debaisieux, L., et al. (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
- Vuilleumier, S., and Bonhoeffer, S. (2015). Contribution of recombination to the evolutionary history of HIV. *Curr. Opin. HIV AIDS* 10, 84–89.

**Conflict of Interest:** 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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hill, Pérez-Losada, Delgado, Benito, Montero, Gil, Sánchez, Cañada-García, García-Bodas, Crandall, Thomson and the Spanish Group for the Study of New HIV Diagnoses. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.