



Spatiotemporal Characteristics of the Largest HIV-1 CRF02_AG Outbreak in Spain: Evidence for Onward Transmissions

*Evangelia-Georgia Kostaki*¹, *Andreas Flampouris*¹, *Timokratis Karamitros*^{2,3}, *Natalia Chueca*⁴, *Marta Alvarez*⁴, *Paz Casas*⁴, *Belen Alejos*⁵, *Angelos Hatzakis*¹, *Federico Garcia*^{4*}, *Dimitrios Paraskevis*¹ and *CoRIS*

¹ Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece, ² Department of Zoology, University of Oxford, Oxford, United Kingdom, ³ Public Health Laboratories, Department of Microbiology, Hellenic Pasteur Institute, Athens, Greece, ⁴ Department of Clinical Microbiology, Hospital Universitario San Cecilio, Instituto de Investigación Ibs, Granada, Spain, ⁵ Centro Nacional de Epidemiología, Universidad de Alcalá de Henares, Madrid, Spain

OPEN ACCESS

Edited by:

Kok Keng Tee,
University of Malaya, Malaysia

Reviewed by:

Fernando Gonzalez-Candelas,
University of Valencia, Spain
Frank van der Meer,
University of Calgary, Canada

*Correspondence:

Federico Garcia
fegarcia@ugr.es

Specialty section:

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

Received: 15 October 2018

Accepted: 12 February 2019

Published: 11 March 2019

Citation:

Kostaki E-G, Flampouris A, Karamitros T, Chueca N, Alvarez M, Casas P, Alejos B, Hatzakis A, Garcia F, Paraskevis D and CoRIS (2019) Spatiotemporal Characteristics of the Largest HIV-1 CRF02_AG Outbreak in Spain: Evidence for Onward Transmissions. *Front. Microbiol.* 10:370. doi: 10.3389/fmicb.2019.00370

Background and Aim: The circulating recombinant form 02_AG (CRF02_AG) is the predominant clade among the human immunodeficiency virus type-1 (HIV-1) non-Bs with a prevalence of 5.97% (95% Confidence Interval-CI: 5.41–6.57%) across Spain. Our aim was to estimate the levels of regional clustering for CRF02_AG and the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain.

Methods: We studied 396 CRF02_AG sequences obtained from HIV-1 diagnosed patients during 2000–2014 from 10 autonomous communities of Spain. Phylogenetic analysis was performed on the 391 CRF02_AG sequences along with all globally sampled CRF02_AG sequences ($N = 3,302$) as references. Phylodynamic and phylogeographic analysis was performed to the largest CRF02_AG monophyletic cluster by a Bayesian method in BEAST v1.8.0 and by reconstructing ancestral states using the criterion of parsimony in Mesquite v3.4, respectively.

Results: The HIV-1 CRF02_AG prevalence differed across Spanish autonomous communities we sampled from ($p < 0.001$). Phylogenetic analysis revealed that 52.7% of the CRF02_AG sequences formed 56 monophyletic clusters, with a range of 2–79 sequences. The CRF02_AG regional dispersal differed across Spain ($p = 0.003$), as suggested by monophyletic clustering. For the largest monophyletic cluster (subepidemic) ($N = 79$), 49.4% of the clustered sequences originated from Madrid, while most sequences (51.9%) had been obtained from men having sex with men (MSM). Molecular clock analysis suggested that the origin (t_{MRCA}) of the CRF02_AG subepidemic was in 2002 (median estimate; 95% Highest Posterior Density-HPD interval: 1999–2004). Additionally, we found significant clustering within the CRF02_AG subepidemic according to the ethnic origin.

Conclusion: CRF02_AG has been introduced as a result of multiple introductions in Spain, following regional dispersal in several cases. We showed that CRF02_AG transmissions were mostly due to regional dispersal in Spain. The hot-spot for the

largest CRF02_AG regional subepidemic in Spain was in Madrid associated with MSM transmission risk group. The existence of subepidemics suggest that several spillovers occurred from Madrid to other areas. CRF02_AG sequences from Hispanics were clustered in a separate subclade suggesting no linkage between the local and Hispanic subepidemics.

Keywords: HIV-1, CRF02_AG, Spain, regional dispersal, spatiotemporal characteristics

INTRODUCTION

Human immunodeficiency virus (HIV)-pandemic has been caused by Group M which is divided into nine subtypes (A, B, C, D, F, G, H, J, and K), six sub-subtypes (A1, A2, A3, A4, F1, F2) but also an extensive list of at least 98 CRFs which result from the recombination of two or more different subtypes (Foley et al., 2016)¹. The distribution of subtypes and CRFs differ greatly across the globe. The majority (46.6%) of the infections worldwide are caused by subtype C, followed by subtype B (12.1%) and subtype A (10.3%); the former is predominant in the Western world (Hemelaar et al., 2018). CRF02_AG is the fourth worldwide (7.7%), however, it accounts for approximately 50% of the HIV-infections in West and Central Africa (Delatorre et al., 2014). CRF01_AE (5.3%), subtype G (4.6%) and D (2.7%) are following (Hemelaar et al., 2018).

The putative origin of CRF02_AG was in Central Africa, thereafter the virus spread into Western Africa establishing a major epidemic (Mir et al., 2016). Further dissemination from Western Africa occurred to Cameroon in the late seventies, to the former Soviet Union in the late nineties and later to Bulgaria and Germany (Mir et al., 2016). CRF02_AG remains one of the most prevalent CRFs in Europe among individuals from highly endemic countries but also in non-migrant populations (Abecasis et al., 2013; Beloukas et al., 2016).

A previous study on 6,633 sequences sampled across Spain, revealed that subtype B was the most prevalent, but the non-B clades were dominated by CRFs. Specifically, it was found that CRF02_AG had a prevalence of 5.97% (95% CI: 5.41–6.57%) and was the most prevalent among the non-B clades (Flampouris et al., 2014). It is estimated that 130,000–160,000 people are living with HIV in Spain (Morán Arribas et al., 2018) while there is a high burden of late diagnosis. Current HIV epidemic in Spain is dominated by infections among MSM (del Amo and Iniesta, 2018).

The aim of the current study was to estimate the levels of regional clustering for CRF02_AG in Spain, to assess parameters associated with its regional dispersal across Spain, but also to estimate the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain. We also investigated whether regional dispersal is due to migration or due to onward transmissions among non-migrants.

Abbreviations: CI, confidence interval; CRFs, circulating recombinant forms; CRF02_AG, circulating recombinant form 02_AG; HIV-1, human immunodeficiency virus type-1; MCMC, Markov chain Monte Carlo; MSM, men having sex with men; PLHIV, people living with HIV; t_{MRC} , time to the most recent common ancestor.

¹<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>

MATERIALS AND METHODS

Study Population

We studied 396 HIV-1 CRF02_AG sequences available in the protease and partial reverse transcriptase regions of the virus genome (*pol gene*). Sequences were obtained from HIV-1 diagnosed patients during 2000–2014 from 10 autonomous communities of Spain. Study population characteristics are shown in **Table 1**. Specifically, patients' samples were merged from two datasets: (a) a representative cohort of HIV-infected patients included in the Research Network on HIV/AIDS (CoRIS) (2004–2013) and (b) Eastern Andalusia Resistance Cohort (2000–2014). A detailed description of the CoRIS cohort has been published previously (Caro-Murillo et al., 2007; Yebra et al., 2012). For this study, fully anonymized nucleotide sequences were retrospectively analyzed thus no written informed consent is required. This research project has been approved by the HU san Cecilio's Ethics Committee.

DNA Sequence Alignment and Phylogenetic Analysis

The patterns of CRF02_AG dispersal in Spain were investigated by means of phylogenetic analysis. Analysis was performed in 391 out of 396 (98.7%) CRF02_AG sequences along with all globally sampled CRF02_AG sequences available on the public HIV-1 sequence database² ($N = 3,302$) as references. A small number of identical sequences (5 of 396, 1.3%) was excluded from analysis to avoid inclusion of duplicates. MEGA v5.2 (Hall, 2013) was used to align sequences using the MUSCLE algorithm and alignments were manually edited according to the encoded reading frame (only codons and no single or double combinations of nucleotides were excluded from the alignment). Codons associated with resistance (IAS mutation list 2017) were excluded to avoid potentially bias on clustering due to convergent evolution at resistance sites (Lewis et al., 2008). The final alignment was consisted of 738 nucleotides. FigTree v1.4 was used for tree visualization and annotation³.

The phylogenetic trees were estimated from the underlying nucleotide sequences. Initially, Maximum likelihood phylogeny reconstruction with bootstrap evaluation was conducted in RAxML v8.0.20 (Stamatakis, 2014) using the general time-reversible (GTR) substitution model and gamma (Γ) distribution. Subsequently, further analysis was performed on the clusters that initially received bootstrap value lower than 75%, by using

²<https://www.hiv.lanl.gov>

³<http://tree.bio.ed.ac.uk/software/figtree/>

a well-justified method based on Bayesian analysis (Kouyos et al., 2010; Paraskevis et al., 2019). In detail, we used the Simple Consensus Maker algorithm⁴ to make the consensus sequence of each cluster with bootstrap value lower than the threshold. The 100 most closely related sequences were found by ordering the consensus sequence with all the CRF02_AG available sequences (391 sequences from our study population and 3,302 reference sequences) according to their similarity, in Mafft v7.4 (Kato et al., 2017). Sequences found within each cluster were then analyzed phylogenetically along with the 100 most closely related sequences to them, using the Bayesian method with the GTR substitution model with Γ distributed rate, as implemented in MrBayes v3.2.2 (Ronquist and Huelsenbeck, 2003). The MCMC ran for 10×10^5 generations (burn-in: 10%), with 4 chains per run, and with MCMC sampling every 1,000 steps. Each MCMC run checked for convergence using Tracer version v1.5 (Drummond et al., 2012). Thereafter, phylogenetic clusters were defined as monophyletic using two different criteria: (i) clusters with bootstrap values greater than 75%, for phylogenetic trees estimated by the Maximum likelihood method, or a posterior probability greater

than 0.82⁵, for phylogenetic trees estimated by the Bayesian method (phylogenetic confidence criterion), and (ii) clusters consisting of at least two sequences sampled from Spain at a proportion greater than 70% compared to total number of sequences within the cluster (geographic criterion). Only phylogenetic clusters fulfilling both criteria were considered as monophyletic.

Phylogenetic Analysis

We further conducted phylogenetic analysis to estimate the spatiotemporal characteristics of the largest CRF02_AG subepidemic (monophyletic cluster consisted of 79 sequences). Analysis was performed only on sequences with available information about their date of sampling (72 of 79, 91.1%). We performed the analysis by a Bayesian method using the HKY nucleotide substitution model, a Γ distributed rate of heterogeneity among sites, an uncorrelated log normal relaxed clock of molecular clock model with TipDates and the birth-death basic reproductive number (R_e) models as implemented in BEAST v1.8.0 (Drummond et al., 2012). Non-informative priors were used for the MCMC runs. MCMC analysis was run for 30×10^6 generations, sampled every 3,000 steps (burn-in: 10%). The MCMC convergence was checked using Tracer v1.5 (Drummond et al., 2012), by estimating the effective sample sizes (ESS > 200). The temporal signal in our data was tested by TempEst (Rambaut et al., 2016).

Phylogeographic Analysis

Phylogeographic analysis was inferred by character reconstruction using the criterion of parsimony on the dated phylogeny using Mesquite v3.4⁶.

Statistical Analysis

Demographic data are summarized using median and interquartile ranges for continuous variables, and absolute and relative frequencies for categorical variables. Statistical analysis for simple comparisons of the relevant distributions across different levels of other categorical variables was carried out using Pearson's chi-squared statistical test. A multivariate logistic regression model was fit to a subset of the original data ($N = 391$), consisting of 304 complete observations. Presence in monophyletic groups was the binary outcome variable, while age, gender, transmission risk group and Spanish autonomous community of sampling were chosen as possible explanatory variables. Statistical analysis was performed on STATA 12-StataCorp LP.

RESULTS

CRF02_AG prevalence was significantly different across the Spanish autonomous communities we sampled from ($p < 0.001$).

⁵All clusters receive posterior probability support >0.82, except for two clusters consisting of two sequences each, for which the posterior probability support was 0.76 and 0.79. The latter were included in the analysis since the confirmatory analysis was based on stringent sampling criteria.

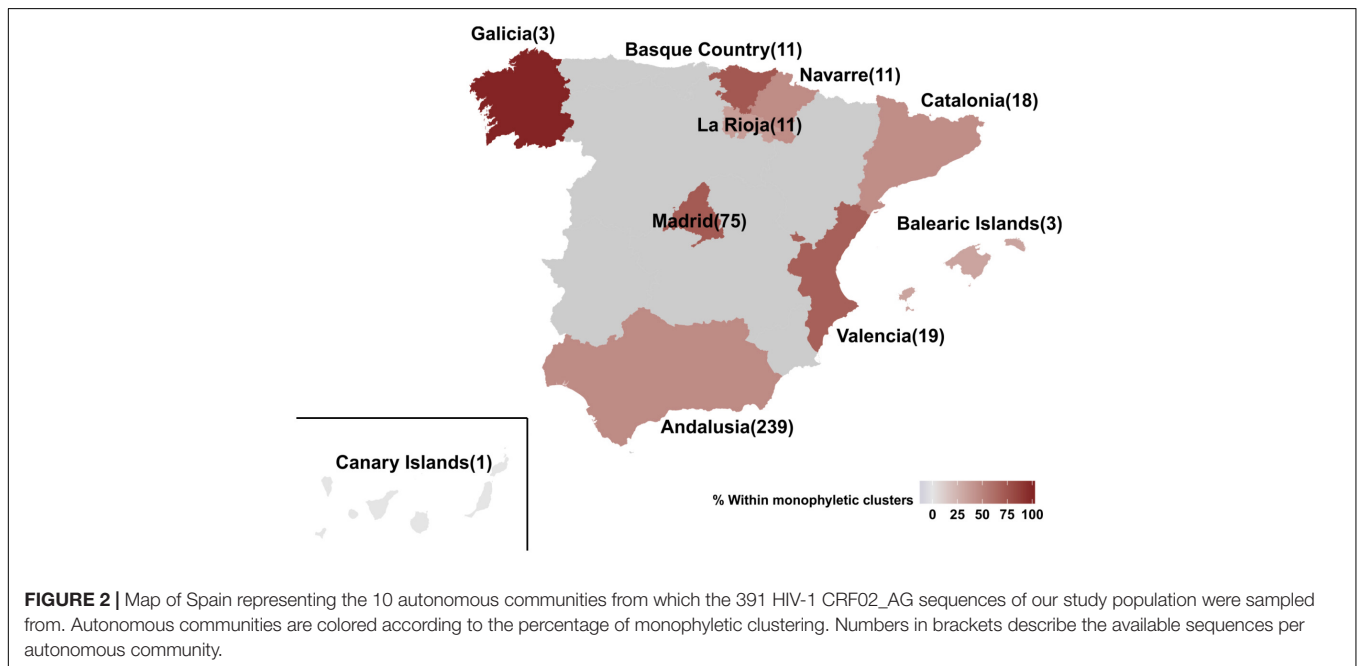
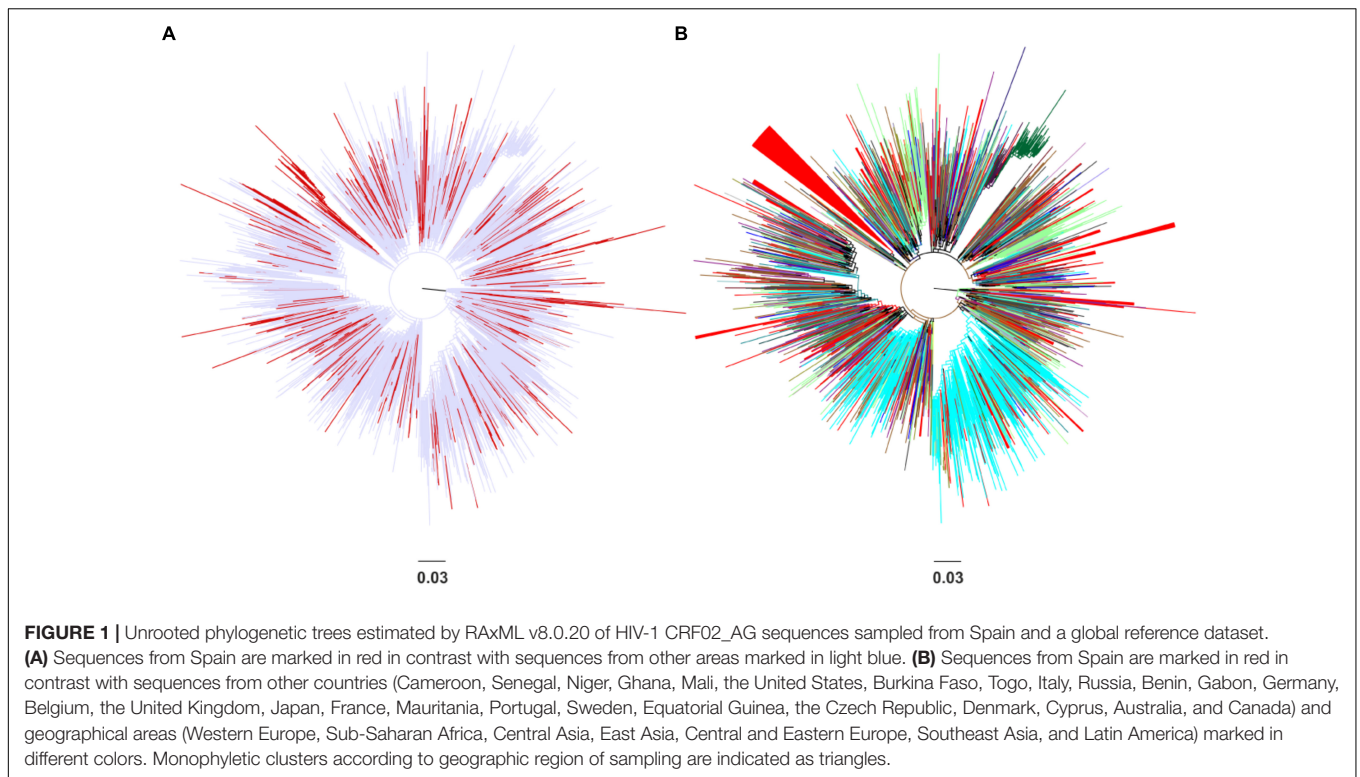
⁶<http://www.mesquiteproject.org>

⁴<https://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html>

TABLE 1 | Characteristics of the study population: (i) total population ($N = 396$), and (ii) a subset of the total population ($N = 391$ of 396) after the exclusion of identical sequences.

Characteristic	Total population	Subset of the total population
Median age [years (IQR ¹)]	35.6 (29.3–43.6) ²	36.2 (29.4–43.6) ³
Gender (N, %)		
Male	233 (58.8)	230 (58.8)
Female	163 (41.2)	161 (41.2)
Transmission risk group (N, %)		
Heterosexuals	158 (39.9)	156 (39.9)
MSM ⁴	62 (15.7)	60 (15.4)
PWID ⁵	1 (0.2)	1 (0.2)
Other/unknown	175 (44.2)	174 (44.5)
Spanish autonomous community of sampling (N, %)		
Andalusia	240 (60.6)	239 (61.3)
Madrid	76 (19.2)	75 (19.2)
Valencia	22 (5.6)	19 (4.9)
Catalonia	18 (4.6)	18 (4.6)
La Rioja	11 (2.8)	11 (2.8)
Navarre	11 (2.8)	11 (2.8)
Basque Country	11 (2.8)	11 (2.8)
Balearic Islands	3 (0.7)	3 (0.7)
Galicia	3 (0.7)	3 (0.7)
Canary Islands	1 (0.2)	1 (0.2)
Total (N, %)	396 (100)	391 (100)

¹IQR, interquartile range; ²N = 313 (of 396 individuals); ³N = 309 (of 391 individuals); ⁴MSM, men having sex with men; ⁵PWID, people who inject drugs.



La Rioja ($N = 11$, 16.7%, 95% CI: 8.6–27.9%), Navarre ($N = 11$, 13.6%, 95% CI: 7–23%), Valencia ($N = 22$, 7.3%, 95% CI: 5.3–12.4%), and Andalusia ($N = 240$, 7%, 95% CI: 6.2–8.0%) were the autonomous communities where CRF02_AG was most prevalent.

Phylogenetic analysis revealed that the CRF02_AG sequences sampled from Spain clustered at different points in the

Maximum-likelihood tree (**Figure 1A**). No specific geographic area was identified as the origin of CRF02_AG in Spain, since the Spanish monophyletic clusters appeared as nested within sequences from different countries (**Figure 1B**). The high-levels of dispersal of the CRF02_AG strains (red branches) across the tree (**Figure 1A**) or the existence of a large number of monophyletic clusters (**Figure 1B**) indicate that the introduction

of CRF02_AG into Spain has been occurred from multiple sources. Specifically, we found that 52.7% (206 of 391) of our study CRF02_AG sequences formed 56 monophyletic clusters, with a range of 2–79 sequences (**Figure 1B**). The median number of sequences per monophyletic cluster was 3 (Interquartile range, IQR: 2–5). Phylogenetic analysis revealed the presence of highly supported phylogenetic clusters consisted of sequences from Spain at proportions >70% (monophyletic clusters). HIV-1 sequences found within the monophyletic clusters suggest a common route of infection. The fact that the majority of sequences have been retrieved from PLHIV residing in Spain implies that monophyletic clusters correspond to transmissions occurred locally (regional transmissions).

The percentage of regional dispersal differed across Spain, as suggested by the monophyletic clustering ($p = 0.003$) (**Figure 2**). Specifically, the highest proportion of regional dispersal was found in Basque Country where 72.7% (8 of 11) of sequences belonged to monophyletic clusters, followed by sequences from Madrid (54 of 75, 72%) and Valencia (13 of 19, 68.4%). Sequences from Andalusia (110 of 239, 46%), Navarre (5 of 11, 45.5%), Catalonia (8 of 18, 44.4%), and La Rioja (4 of 11, 36.4%) showed the lowest monophyly levels. For Galicia ($N = 3$), Canary Islands ($N = 1$), and Balearic Islands ($N = 3$) less than 10 sequences were available. The current study was performed using data from the representative cohort CoRIS (Caro-Murillo et al., 2007; Yebra et al., 2012; Vourli et al., 2019) and the Eastern Andalusia Resistance Cohort. The study population included sequences sampled during 2000–2014 from 10 autonomous communities of Spain, suggesting high geographic coverage. The proportion of sequences within each cluster was comparable across different communities, except for Andalusia which was overrepresented. Despite its over-representativeness, Andalusia was found among the communities with the lowest levels of monophyly.

Using a multivariate logistic regression model, we found that the MSM transmission risk group was positively associated with the regional clustering ($p < 0.001$), having adjusted for the rest of the incorporated variables (**Table 2**). This association was mostly due to the large monophyletic cluster in which the majority of PLHIV were MSM.

The largest CRF02_AG monophyletic cluster (subepidemic) consisted of 79 sequences (**Figure 1B**). We found that 49.4% (39 of 79) of the clustered sequences originated from Madrid and most sequences (41 of 79, 51.9%) had been obtained from MSM. Phylogenetic analysis also revealed the existence of two nested clusters from Japan ($N = 7$, 8.9%) and Sweden ($N = 3$, 3.8%). The nested clusters consisted of reference sequences that were available in the HIV-1 sequence database.

To describe the temporal patterns of the largest CRF02_AG regional epidemic in Spain, we carried out phylodynamic analysis. There was evidence for temporal signal in the sequences found within the largest CRF02_AG monophyletic cluster, as tested by the TempEst program ($R = 0.721$). Moreover, the temporal structure in this data was assessed by estimating the ESS for all parameters of the MCMC analysis. In the MCMC analysis the ESS for all parameters were $>> 200$. As explained in detail in the methods section, phylodynamic analysis was performed only on sequences with available information about their date

TABLE 2 | Multivariate logistic regression estimates using the presence in monophyletic groups as the binary outcome variable.

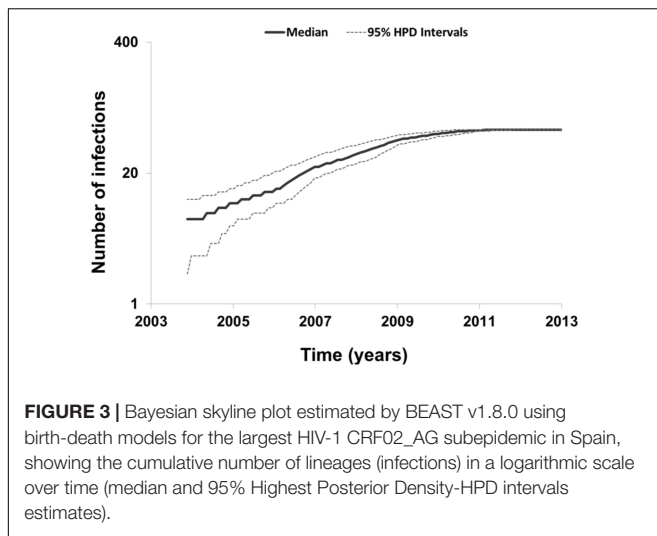
Explanatory variable	Odds ratio	95% Conf. interval	P-value
Age	1.01	(0.99, 1.03)	0.506
Gender (*Male)			
Female	1.09	(0.65, 1.84)	0.732
Transmission risk group (*Heterosexuals)			
MSM ¹	6.66	(2.57, 17.23)	<0.001
PWID ²	1	–	–
Other/unknown	0.63	(0.34–1.18)	0.153
Spanish autonomous community of sampling (*Andalusia)			
Madrid	0.88	(0.41, 1.91)	0.749
Valencia	1.04	(0.33, 3.22)	0.948
Catalonia	0.45	(0.15, 1.34)	0.152
La Rioja	0.48	(0.12, 1.90)	0.299
Navarre	0.47	(0.12, 1.82)	0.277
Basque Country	1.19	(0.26, 5.29)	0.823
Balearic Islands	0.33	(0.03, 3.93)	0.381
Galicia	1	–	–
Canary Islands	1	–	–

The model was fit to a subset of the original data ($N = 391$), consisting of 304 complete observations. *Reference category: ¹MSM, men having sex with men; ²PWID, people who inject drugs.

of sampling ($N = 72$). For 57 out of 72 (79.2%) sequences found within the largest CRF02_AG monophyletic cluster, the sampling area was in Spain (**Table 3**). Molecular clock analysis suggested that the time to most recent common ancestor (t_{MRCA}) of the CRF02_AG subepidemic was in 2002 (median estimate; 95% Highest Posterior Density-HPD interval: 1999–2004). The t_{MRCA} should be considered as the approximate time of infection of the potential founder of the CRF02_AG subepidemic in Spain sampled in our data. The Bayesian skyline plot showed that the CRF02_AG subepidemic growth occurred during a period of 8 years and specifically between 2003 and 2011 (**Figure 3**).

TABLE 3 | Sampling region/country for sequences with known date of sampling found within the largest CRF02_AG monophyletic cluster (subepidemic) in Spain.

Region of sampling	Country of sampling	Number of sequences per country (%)	Number of sequences per region (%)
Western Europe	Spain	57 (79.2)	65 (90.3)
	Switzerland	4 (5.6)	
	Sweden	3 (4.2)	
	Germany	1 (1.2)	
	Japan	7 (9.8)	
East Asia	Japan	7 (9.8)	7 (9.7)
Total		72 (100)	72 (100)

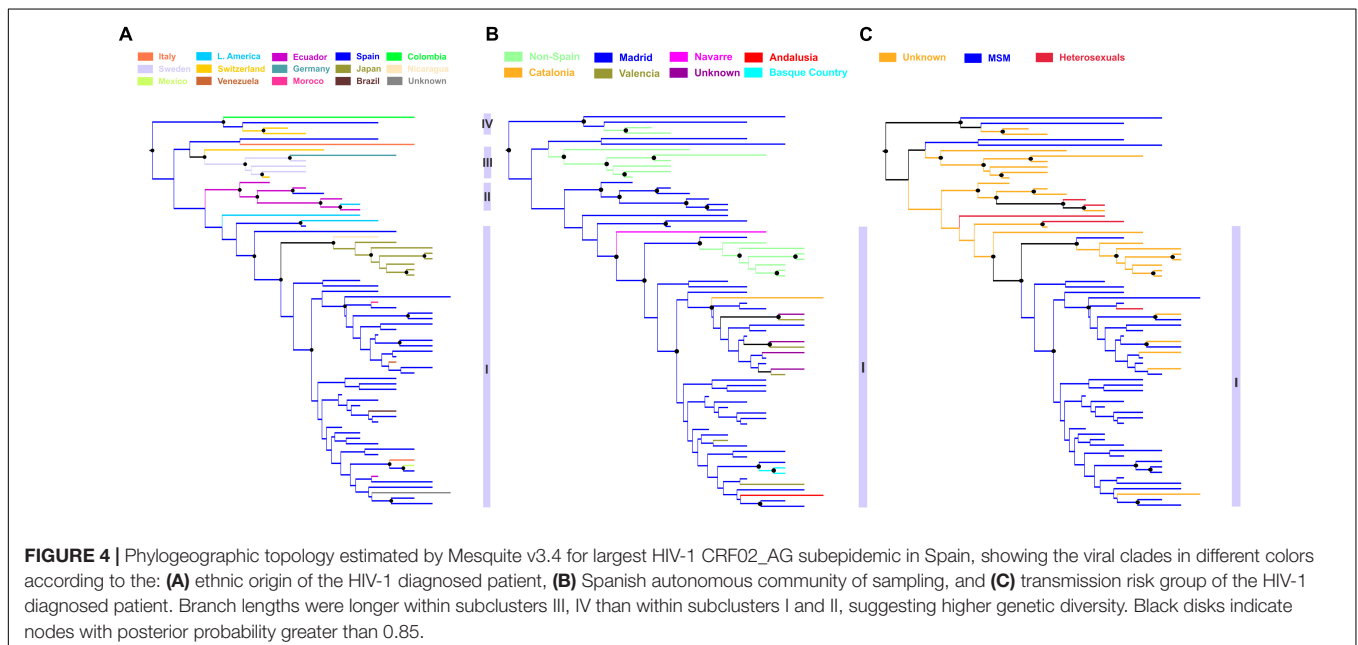


Phylogeographic analysis revealed significant clustering within the CRF02_AG subepidemic according to the ethnic origin (**Figure 4A**). Specifically, four distinct subclades were found. The subclade division was based on high posterior probability support (>0.85) and epidemiologic criteria (i.e., the ethnic origin of PLHIV within each cluster). The first subclade was consisted of sequences from Spain ($N = 44$) including as nested the cluster from Japan ($N = 7$) (subclade I) (**Figures 4A,B**). The majority of Spanish lineages were from individuals living in Madrid ($N = 30$, 68.2%) reported MSM ($N = 37$, 84.1%) as transmission risk group (**Figures 4B,C**). Except from Madrid, Spanish sequences were obtained across five different autonomous communities of Spain (**Figure 4B**). The second subclade was consisted of sequences from Hispanics from Spain ($N = 1$, 17%) and Latin America ($N = 5$, 83%)

(subclade II) (**Figure 4A**). The third subclade consisted of sequences ($N = 6$) originating from Western Europe (Sweden, Switzerland, and Germany) (subclade III) and the fourth of sequences ($N = 4$) from Colombia, Spain, and Switzerland (subclade IV) (**Figure 4A**). Subclades III and IV were more genetically divergent than the two others (as suggested from the within-subcluster branch lengths) (**Figure 4A**). Notably, the geographic origin of the CRF02_AG clade including sequences from Spain (subclades I and II) was estimated in Madrid, from where it spread to other autonomous communities within the country as well as outside Spain (**Figures 4A,B**).

DISCUSSION

In this study we used a collection of sequences sampled from 10 Spanish autonomous communities and a combination of phylogenetic, phylogeographic and phylodynamic analyses, in order to investigate the characteristics of the HIV-1 CRF02_AG epidemic in this country. We further shed light on the spatiotemporal characteristics of the largest CRF02_AG subepidemic in Spain. Our analysis suggests that CRF02_AG has been introduced as a result of multiple introductions in Spain, following regional dispersal in several cases. However, we found considerable variation in the patterns of regional clustering in Spain with Basque area, Madrid and Valencia to be the areas with the higher proportion of regional transmissions. We estimated that the epidemic growth of the largest monophyletic cluster continued until 2011. The hub for the largest cluster was in Madrid from where cross-border and local transmissions to other autonomous communities were monitored. Notably, CRF02_AG sequences within the largest cluster in Madrid were detected as far as from Japan, comprising a local epidemic, as well as from Latin America. Although an alternative hypothesis cannot be



entirely excluded, our suggestion that the Japanese local epidemic originated from Spain provides the most plausible hypothesis given the existing sampling. The long-distance dispersal of CRF02_AG is not strange since some of the sequences were obtained from Hispanic populations outside Spain, a finding that can be reconciled because of the historical links between Spain and Hispanic populations of the Americas. On the other hand, the high popularity of Spain as a travel destination could explain the outgoing transmissions to Japan. CRF02_AG regional dispersal in Spain was associated with MSM; a finding that was probably due to the pattern found for the largest cluster in Madrid.

The identification of Madrid as a hub for the largest cluster of CRF02_AG in Spain and that only small clusters were found outside Madrid, suggest that transmission networking is probably more extensive in the capital than in peripheral cities. Notably, similar findings were reported for Paris and Rome which provided as the source for the CRF02_AG transmissions across France and Italy, respectively (Giuliani et al., 2013; Chaillon et al., 2017). The identity of capital cities as hubs for within country dispersal could be explained by the fact that they provide social, economic cultural and recreational centers as well as major hubs for international traveling. These findings could be useful for public health stakeholders to focus their preventive actions in large cities in Europe.

The prevalence of CRF02_AG has been estimated to 3.0% in Western and Central Europe and North America, being the most frequently circulated among the CRFs (Hemelaar et al., 2018). Although sporadic CRF02_AG cases have been increasingly detected in Europe and North America as a result of population mobility from endemic regions, regional dispersal was also found in several areas in Europe including MSM and heterosexual populations in France, Belgium, Switzerland, and Spain (Dauwe et al., 2015), as well as people who inject drugs in Bulgaria (Alexiev et al., 2016). Specifically, only one small cluster was found in Belgium, while 28% of the CRF02_AG sequences belonged to Swiss-specific subclusters (Von Wyl et al., 2011). In France, 20.4% of the PLHIV primarily infected with CRF02_AG belonged to transmission networks based on analysis using a genetic distance cut-off (Chaillon et al., 2017). This proportion was the highest for all HIV-1 subtypes (Chaillon et al., 2017). A large monophyletic cluster of CRF02_AG was also monitored among PLHIV with primary infection in Southeastern France (Tamalet et al., 2015). Similarly, the proportion of CRF02_AG sequences within monophyletic clusters was reported: 29.8, 27.3, 25.4, and 18.2% for Germany, Norway, Austria, and Italy, respectively (Paraskevis et al., 2019). These findings suggest a common pattern of CRF02_AG dispersal in Europe, where although CRF02_AG have been associated with highly endemic areas (Abecasis et al., 2013) as their putative source, a considerable proportion of their sequences fall within monophyletic clusters. It remains unclear, however, whether onward transmissions in Europe occur among migrant or non-migrant populations.

In the current study using a state-of-the-art molecular epidemiology approach, we were able to describe the

spatiotemporal characteristics of the CRF02_AG infection in Spain and to investigate the nature of onward transmissions across different ethnic groups. Notably, we found that the majority of CRF02_AG sequences within the largest cluster were retrieved from non-migrant populations, suggesting that the within the country infections are due to onward transmissions among non-migrants. To our knowledge this is one of the few studies providing evidence that local dispersal of non-B subtypes is associated with non-migrants. These findings can be important for public health and particularly for the design of targeted interventions for the populations at the higher risk.

CENTERS AND INVESTIGATORS INVOLVED IN CoRIS

Executive Committee

Santiago Moreno, Inma Jarrín, David Dalmau, Maria Luisa Navarro, Maria Isabel González, Jose Luis Blanco, Federico Garcia, Rafael Rubio, Jose Antonio Iribarren, Félix Gutiérrez, Francesc Vidal, Juan Berenguer, Juan González.

Fieldwork, data management and analysis: Belén Alejos, Victoria Hernando, Cristina Moreno, Carlos Iniesta, Luis Miguel Garcia Sousa, Nieves Sanz Perez.

BioBank HIV

Hospital General Universitario Gregorio Marañón: M Ángeles Muñoz-Fernández, Isabel María García-Merino, Irene Consuegra Fernández, Coral Gómez Rico, Jorge Gallego de la Fuente, Paula Palau Concejo.

Participating Centres

Hospital General Universitario de Alicante (Alicante):

Joaquín Portilla, Esperanza Merino, Sergio Reus, Vicente Boix, Livia Giner, Carmen Gadea, Irene Portilla, María Pampliega, Marcos Díez, Juan Carlos Rodríguez, José Sánchez-Payá.

Hospital Universitario de Canarias (San Cristobal de la Laguna): Juan Luis Gómez, Jehovana Hernández, María Remedios Alemán, María del Mar Alonso, María Inmaculada Hernández, Felicitas Díaz-Flores, Dácil García, Ricardo Pelazas, Ana López Lirola.

Hospital Universitario Central de Asturias (Oviedo): José Sanz Moreno, Alberto Arranz Caso, Cristina Hernández Gutiérrez, María Novella Mena.

Hospital Universitario 12 de Octubre (Madrid): Rafael Rubio, Federico Pulido, Otilia Bisbal, Asunción Hernando, Lourdes Domínguez, David Rial Crestelo, Laura Bermejo, Mireia Santacreu.

Hospital Universitario de Donostia (Donostia-San Sebastián): José Antonio Iribarren, Julio Arrizabalaga, María José Aramburu, Xabier Camino, Francisco Rodríguez-Arondo, Miguel Ángel von Wichmann, Lidia Pascual Tomé, Miguel Ángel Goenaga, M^a Jesús Bustinduy, Harkaitz Azkune, Maialen Iburguren, Aitziber Lizardi, Xabier Kortajarena.

Hospital General Universitario De Elche (Elche): Félix Gutiérrez, Mar Masiá, Sergio Padilla, Andrés Navarro, Fernando

Montolio, Catalina Robledano, Joan Gregori Colomé, Araceli Adsuar, Rafael Pascual, Marta Fernández, Elena García, José Alberto García, Xavier Barber.

Hospital General Universitario Gregorio Marañón (Madrid): Juan Berenguer, Juan Carlos López Bernaldo de Quirós, Isabel Gutiérrez, Margarita Ramírez, Belén Padilla, Paloma Gijón, Teresa Aldamiz-Echevarría, Francisco Tejerina, Francisco José Parras, Pascual Balsalobre, Cristina Diez, Leire Pérez Latorre.

Hospital Universitari de Tarragona Joan XXIII (Tarragona): Francesc Vidal, Joaquín Peraire, Consuelo Viladés, Sergio Veloso, Montserrat Vargas, Miguel López-Dupla, Montserrat Olona, Anna Rull, Esther Rodríguez-Gallego, Verónica Alba.

Hospital Universitario y Politécnico de La Fe (Valencia): Marta Montero Alonso, José López Aldegue, Marino Blanes Juliá, María Tasia Pitarch, Iván Castro Hernández, Eva Calabuig Muñoz, Sandra Cuéllar Tovar, Miguel Salavert Lletí, Juan Fernández Navarro, Jose Miguel Molina.

Hospital Universitario La Paz/IdiPAZ: Juan González-garcía, Francisco Arnalich, José Ramón Arribas, Jose Ignacio Bernardino de la Serna, Juan Miguel Castro, Luis Escosa, Pedro Herranz, Victor Hontañón, Silvia García-Bujalance, Milagros García López-Hortelano, Alicia González-Baeza, Maria Luz Martín-Carbonero, Mario Mayoral, Maria Jose Mellado, Rafael Esteban Micán, Rocio Montejano, María Luisa Montes, Victoria Moreno, Ignacio Pérez-Valero, Berta Rodés, Talia Sainz, Elena Sendagorta, Natalia Stella Alcáriz, Eulalia Valencia.

Hospital San Pedro Centro de Investigación Biomédica de La Rioja (CIBIR) (Logroño): José Ramón Blanco, José Antonio Oteo, Valvanera Ibarra, Luis Metola, Mercedes Sanz, Laura Pérez-Martínez.

Hospital Universitari MutuaTerrassa (Terrasa): David Dalmau, Angels Jaén, Montse Sanmartí, Mireia Cairó, Javier Martinez-Lacasa, Pablo Velli, Roser Font, Mariona Xercavins, Noemí Alonso.

Complejo Hospitalario de Navarra (Pamplona) María Rivero, Jesús Repáraz, María Gracia Ruiz de Alda, María Teresa de León Cano, Beatriz Pierola Ruiz de Galarreta.

Hospital Universitario de La Princesa (Madrid): Ignacio de los Santos, Jesús Sanz Sanz, Ana Salas Aparicio, Cristina Sarriá Cepeda, Lucio Garcia-Fraile Fraile, Enrique Martín Gayo.

Hospital Universitario Ramón y Cajal (Madrid): Santiago Moreno, José Luis Casado, Fernando Dronza, Ana Moreno, María Jesús Pérez Elías, Cristina Gómez Ayerbe, Carolina Gutiérrez, Nadia Madrid, Santos del Campo Terrón, Paloma Martí, Uxua Ansa, Sergio Serrano, María Jesús Vivancos.

Hospital General Universitario Reina Sofía (Murcia): Enrique Bernal, Alfredo Cano, Antonia Alcaraz García, Joaquín Bravo Urbietta, Ángeles Muñoz, Maria Jose Alcaraz, Maria del Carmen Villalba.

Hospital Universitario San Cecilio (Granada): Federico García, José Hernández, Alejandro Peña, Leopoldo Muñoz, Paz Casas, Marta Alvarez, Natalia Chueca, David Vinuesa, Clara Martínez-Montes, Fernando García, Carlos Guerrero-Beltran.

Centro Sanitario Sandoval (Madrid): Jorge Del Romero, Carmen Rodríguez, Teresa Puerta, Juan Carlos Carrió, Mar Vera, Juan Ballesteros, Oskar Ayerdi.

Hospital Clínico Universitario de Santiago (Santiago de Compostela): Antonio Antela, Elena Losada, Antonio Aguilera.

Hospital Universitario Son Espases (Palma de Mallorca): Melchor Riera, María Peñaranda, María Leyes, M^a Angels Ribas, Antoni A Campins, Carmen Vidal, Francisco Fanjul, Javier Murillas, Francisco Homar.

Hospital Universitario Virgen de la Victoria (Málaga): Jesús Santos, Crisitina Gómez Ayerbe, Isabel Viciano, Rosario Palacios, Carmen María González.

Hospital Universitario Virgen del Rocío (Sevilla): Pompeyo Viciano, Nuria Espinosa, Luis Fernando López-Cortés.

Hospital Universitario de Bellvitge (Hospitalet de Llobregat): Daniel Podzamczar, Elena Ferrer, Arkaitz Imaz, Juan Tiraboschi, Ana Silva, María Saumoy.

Hospital Costa del Sol (Marbella): Julián Olalla, Alfonso del Arco, Javier de la torre, José Luis Prada, José María García de Lomas Guerrero, Javier Pérez Stachowski.

Hospital Marina Baixa (VilaJollosa, Alicante): Concepción Amador.

Hospital General Universitario Santa Lucía (Cartagena): Onofre Juan Martínez, Francisco Jesús Vera, Lorena Martínez, Josefina García, Begoña Alcaraz, Amaya Jimeno.

Complejo Hospitalario Universitario a Coruña (Chuc) (A Coruña): Angeles Castro Iglesias, Berta Pernas Souto, Alvaro Mena de Cea.

Hospital Universitario Virgen de la Arrixaca (El Palmar): Carlos Galera, Helena Albendin, Aurora Pérez, Asunción Iborra, Antonio Moreno, Maria Angustias Merlos, Asunción Vidal.

Hospital Universitario Infanta Sofía (San Sebastian de los Reyes): Inés Suárez-García, Eduardo Malmierca, Patricia González-Ruano, Dolores Martín Rodrigo.

Complejo Hospitalario de Jaén (Jaén): Mohamed Omar Mohamed-Balghata, María Amparo Gómez Vidal. Nuestra Señora de Valme: Juan A Pineda, Juan Macías, Samuel Bernal. Unidad de Biología y Variabilidad del VIH.

Centro Nacional de Microbiología, ISCIII: Miguel Thomson, Elena Delgado, Sonia Benito, Vanessa Montero.

DATA AVAILABILITY

The HIV-1 nucleotide sequences of our study population are available in GenBank with accession numbers MK588005–MK588395.

AUTHOR CONTRIBUTIONS

E-GK did the analysis and prepared the figures and the manuscript. AF and TK organized the data and built a database. NC, MA, PC, BA, and FG participated in the study design, performed the data collection across Spain, and provided critical comments about the manuscript. AH contributed to the study design and provided critical comments about the manuscript. DP performed the study design and supervised, and contributed to manuscript writing and editing.

FUNDING

This work was supported by Plan Nacional de I+D+I and Fondo Europeo de Desarrollo Regional-FEDER (RD16/0025/0040 and

PI18/00819) and Fundación Progreso y Salud, Junta de Andalucía (PI-0550-2017). The study was in part funded by the Hellenic Scientific Society for the Study of AIDS and Sexually Transmitted Diseases.

REFERENCES

- Abecasis, A. B., Wensing, A. M. J., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A., et al. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 10:7. doi: 10.1186/1742-4690-10-7
- Alexiev, I., Shankar, A., Dimitrova, R., Gancheva, A., Kostadinova, A., Teoharov, P., et al. (2016). Origin and spread of HIV-1 in persons who inject drugs in Bulgaria. *Infect. Genet. Evol.* 46, 269–278. doi: 10.1016/j.meegid.2016.05.029
- Beloukas, A., Psarris, A., Giannelou, P., Kostaki, E., Hatzakis, A., and Paraskevis, D. (2016). Molecular epidemiology of HIV-1 infection in Europe: an overview. *Infect. Genet. Evol.* 46, 180–189. doi: 10.1016/j.meegid.2016.06.033
- Caro-Murillo, A. M., Castilla, J., Perez-Hoyos, S., Miro, J. M., Podzamczar, D., Rubio, R., et al. (2007). Spanish cohort of naïve HIV-infected patients (CoRIS): rationale, organization and initial results | Cohorte RIS de pacientes con infección por VIH sin tratamiento antirretroviral previo (CoRIS): Metodología y primeros resultados. *Enferm. Infecc. Microbiol. Clin.* 25, 23–31. doi: 10.1157/13096749
- Chaillon, A., Essat, A., Frange, P., Smith, D. M., Delaugerre, C., Barin, F., et al. (2017). Spatiotemporal dynamics of HIV-1 transmission in France (1999–2014) and impact of targeted prevention strategies. *Retrovirology* 14, 1–12. doi: 10.1186/s12977-017-0339-4
- Dauwe, K., Mortier, V., Schaulvliege, M., Van Den Heuvel, A., Franssen, K., Servais, J. Y., et al. (2015). Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate. *BMC Infect. Dis.* 15:524. doi: 10.1186/s12879-015-1217-0
- del Amo, J., and Iñesta, C. (2018). Missed opportunities within the health system for the diagnosis of HIV infection in MSM in Spain: greater commitment and action is required. *Enferm. Infecc. Microbiol. Clin.* 36, 463–464. doi: 10.1016/j.eimc.2018.06.003
- Delatorre, E., Mir, D., and Bello, G. (2014). Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa. *PLoS One* 9:e98908. doi: 10.1371/journal.pone.0098908
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075
- Flampouris, A., Alvarez, M., Kostaki, V., Chueca, N., Monge, S., García-Bujalance, S., et al. (2014). “Molecular epidemiology analysis of HIV-1 epidemic in Spain,” in *VI Congreso Nacional de GeSIDA*, ed. B. A. Grajera (Malaga: Elsevier).
- Foley, B. T., Leitner, T., Paraskevis, D., and Peeters, M. (2016). Primate immunodeficiency virus classification and nomenclature: review. *Infect. Genet. Evol.* 46, 150–158. doi: 10.1016/j.meegid.2016.10.018
- Giuliani, M., Santoro, M. M., Lo Presti, A., Cella, E., Scognamiglio, P., Lai, A., et al. (2013). Circulation of HIV-1 CRF02-AG among MSM population in Central Italy: a molecular epidemiology-based study. *Biomed. Res. Int.* 2013:810617. doi: 10.1155/2013/810617
- Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* 30, 1229–1235. doi: 10.1093/molbev/mst012
- Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., et al. (2018). Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* 19, 143–155. doi: 10.1016/S1473-3099(18)30647-9
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* doi: 10.1093/bib/bbx108 [Epub ahead of print].
- Kouyos, R. D., von Wyl, V., Yerly, S., Böni, J., Taffé, P., Shah, C., et al. (2010). Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J. Infect. Dis.* 201, 1488–1497. doi: 10.1086/651951
- Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Leigh Brown, A. J. (2008). Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 5:e50. doi: 10.1371/journal.pmed.0050050
- Mir, D., Jung, M., Delatorre, E., Vidal, N., Peeters, M., and Bello, G. (2016). Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. *Infect. Genet. Evol.* 46, 190–199. doi: 10.1016/j.meegid.2016.05.017
- Morán Arribas, M., Rivero, A., Fernández, E., Poveda, T., and Caylá, J. A. (2018). Burden of HIV infection, vulnerable populations and access barriers to healthcare. *Enferm. Infecc. Microbiol. Clin.* 36(Suppl. 1), 3–9. doi: 10.1016/S0213-005X(18)30239-8
- Paraskevis, D., Beloukas, A., Stasinou, K., Pantazis, N., de Mendoza, C., Bannert, N., et al. (2019). HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. *BMC Med.* 17:4. doi: 10.1186/s12916-018-1241-1
- Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vev007. doi: 10.1093/ve/vev007
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Tamalet, C., Ravaux, I., Moreau, J., Bregigeon, S., Tourres, C., Richet, H., et al. (2015). Emergence of clusters of CRF02_AG and B human immunodeficiency viral strains among men having sex with men exhibiting HIV primary infection in Southeastern France. *J. Med. Virol.* 87, 1327–1333. doi: 10.1002/jmv.24184
- Von Wyl, V., Kouyos, R. D., Yerly, S., Böni, J., Shah, C., Bürgisser, P., et al. (2011). The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *J. Infect. Dis.* 204, 1095–1103. doi: 10.1093/infdis/jir491
- Vourli, G., Pharris, A., Cazein, F., Costagliola, D., Dabis, F., Del Amo, J., et al. (2019). Are European HIV cohort data within EuroCoord representative of the diagnosed HIV population? *Aids* 33, 133–143. doi: 10.1097/QAD.0000000000002034
- Yebra, G., De Mulder, M., Martín, L., Rodríguez, C., Labarga, P., Viciano, I., et al. (2012). Most HIV type 1 non-B infections in the Spanish cohort of antiretroviral treatment-naïve HIV-infected Patients (CoRIS) are due to recombinant viruses. *J. Clin. Microbiol.* 50, 407–413. doi: 10.1128/JCM.05798-11

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

Copyright © 2019 Kostaki, Flampouris, Karamitros, Chueca, Alvarez, Casas, Alejos, Hatzakis, Garcia, Paraskevis and CoRIS. 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.