Origin and Population Dynamics of a Novel HIV-1 Subtype G Clade Circulating in Cape Verde and Portugal

Isabel Inês M. de Pina-Araujo[1☯], Edson Delatorre[2☯], Monick L. Guimarães[2], Mariza G. Morgado[2], Gonzalo Bello[2]*

1 Departamento de Ciência e Tecnologia, Universidade de Cabo Verde, Praia, Santiago, Cabo Verde,
2 Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

☯ These authors contributed equally to this work.

* gbellobr@gmail.com

Abstract

The human immunodeficiency virus type 1 (HIV-1) subtype G is the most prevalent and second most prevalent HIV-1 clade in Cape Verde and Portugal, respectively; but there is no information about the origin and spatiotemporal dispersal pattern of this HIV-1 clade circulating in those countries. To this end, we used Maximum Likelihood and Bayesian coalescent-based methods to analyze a collection of 578 HIV-1 subtype G *pol* sequences sampled throughout Portugal, Cape Verde and 11 other countries from West and Central Africa over a period of 22 years (1992 to 2013). Our analyses indicate that most subtype G sequences from Cape Verde (80%) and Portugal (95%) branched together in a distinct monophyletic cluster (here called G_{CV-PT}). The G_{CV-PT} clade probably emerged after a single migration of the virus out of Central Africa into Cape Verde between the late 1970s and the middle 1980s, followed by a rapid dissemination to Portugal a couple of years later. Reconstruction of the demographic history of the G_{CV-PT} clade circulating in Cape Verde and Portugal indicates that this viral clade displayed an initial phase of exponential growth during the 1980s and 1990s, followed by a decline in growth rate since the early 2000s. Our data also indicate that during the exponential growth phase the G_{CV-PT} clade recombined with a preexisting subtype B viral strain circulating in Portugal, originating the CRF14_BG clade that was later disseminated to Spain and Cape Verde. Historical and recent human population movements between Angola, Cape Verde and Portugal probably played a key role in the origin and dispersal of the G_{CV-PT} and CRF14_BG clades.

Introduction

The global dissemination of human immunodeficiency virus type 1 (HIV-1) group M, the pandemic clade of HIV, resulted from the random exportation out of Central Africa of a few viral strains designated as subtypes (A–D, F–H, J and K) and inter-subtype circulating recombinant forms (CRFs) [1].

Subtype G is the sixth most prevalent HIV-1 clade in the world accounting for nearly 5% of all global infections [2]. This subtype reaches the highest prevalence in some African countries, comprising 30–50% of HIV-1 infections in Cape Verde [3,4] and Nigeria [5–12], and 5–15% of HIV-1 infections in Angola [13–15], Benin [16], Niger [17,18] and Togo [19,20]. A recent study conducted by our group suggests that subtype G most probably emerged in Central Africa around the late 1960s and was rapidly disseminated into the West and West Central African regions [21]. This study showed that basal subtype G lineages (G_{CA}) were mostly restricted to Central and West Central African countries. Two subtype G strains, however, gained access to West Africa between the middle and the late 1970s and fueled secondary local outbreaks, leading to the origin of two major subtype G West African clades (G_{WA-I} and G_{WA-II}).

Some subtype G strains where also disseminated out of the African continent and the most remarkable example is Portugal where subtype G is the second most prevalent HIV-1 clade (>10%), after subtype B (> 40%) [22–25]. The high prevalence of subtypes B and G in Portugal has also promoted the appearance of different types of B/G recombinant strains, including one circulating recombinant form (CRF14_BG) that was initially identified in Galicia, Northern Spain [26]. According to a previous study, the CRF14_BG probably emerged in Portugal in the early 1990s and later spread to Galicia in the late 1990s as a consequence of the mobility of HIV-infected injecting drug users (IDUs) [27].

Notably, about two-thirds of the subtype G viruses previously described in Portugal were found in individuals from Angola and Cape Verde [23]. These countries are two former Portuguese African colonies that have strong historic links and maintain ongoing relationships with Portugal and displayed a relatively high prevalence of subtype G [3,4,13–15]. The high numbers of immigrants from Angola and Cape Verde who enter Portugal and also those Portuguese returning after living in the former Portuguese African colonies from the 1970s onwards [28], supposes a potential risk for introduction of HIV-1 subtype G strains in Portugal. However, the precise evolutionary relationship between subtype G viruses circulating in Angola, Cape Verde and Portugal remains unknown.

The objective of this study was to reconstruct the phylogenetic relationship, onset date and dissemination routes of the HIV-1 subtype G clades circulating in Angola, Cape Verde and Portugal. To this end, we used Maximum Likelihood and Bayesian coalescent-based approaches to analyze 578 HIV-1 subtype G *pol* sequences isolated from Portugal, West Africa and Central Africa over a period of 22 years (1992 to 2013).

Materials and Methods

Sequence dataset

All HIV-1 subtype G *pol* sequences from Portugal and West/Central African countries that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nt 2253–3272 relative to HXB2 clone) and for which the sampling year was known, were downloaded from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by December 2014. The same procedure was adopted to obtain the subtype G *pol* fragment of all CRF14_BG sequences from Portugal and Spain (the main countries were this CRF circulates) with full-length genome characterization up to date. The subtype assignment of all sequences was confirmed by REGA HIV subtyping tool v.2 [29] and bootscanning analysis. In bootscanning analyses, supporting branching of query sequences with HIV-1 group M subtype reference sequences was determined in Neighbor-Joining trees constructed with the Kimura two-parameter model, within a 250bp window moving in steps of 10 bases, using Simplot software v.3.5.1 [30]. Sequences with incorrect classification, multiple sequences from the same individual and sequences from countries poorly represented ($n < 4$ sequences) were removed, resulting in a final data set of

Table 1. HIV-1 subtype G *pol* dataset.

Los Alamos classification	Region	Country	N	Sampling interval
Subtype G	West Africa	Benin	15	2004–2009
		Ghana	9	2002–2010
		Cape Verde	60	2005–2011
		Nigeria	223	1992–2013
		Senegal	12	1998–2010
		Togo	28	2006–2011
	West Central Africa	Cameroon	62	1997–2012
		Gabon	6	2000–2008
		Equatorial Guinea	4	2005–2009
	Central Africa	Angola	20	1997–2010
		DRC	12	1993–2007
		Republic of Congo	8	2003
	Europe	Portugal	107	1998–2008
CRF14_BG	Europe	Portugal	2[a]	1998–2008
		Spain	10[a]	1999–2005

[a] Subtype G *pol* fragments recovered from full-length CRF14_BG reference sequences.

doi:10.1371/journal.pone.0127384.t001

578 HIV-1 subtype G *pol* sequences (Table 1). Sequence's GenBank accession numbers are available in S1 Table. All codon positions known to be associated with major antiretroviral drug resistance were maintained in the final alignment because phylogenetic trees constructed on alignments with or without such positions resulted in the same overall topology (data not shown). The presence of phylogenetic signal and substitution saturation in our datasets was investigated by: 1) using the likelihood mapping analysis [31] performed with TREE-PUZZLE v5.2 program [32] implemented in the online web platform Mobyle@Pasteur v1.5 [33], and 2) plotting the observed number of transitions and transversions against genetic distance for each pairwise comparison, calculated under the GTR+I+G nucleotide substitution model using DAMBE v5.3 program [34].

Identification of major HIV-1 subtype G clades

Major HIV-1 subtype G clades were identified by Maximum Likelihood (ML) phylogenetic analysis. A ML phylogeny was constructed with the PhyML 3.0 program [35] using an online web server [36]. The ML tree was inferred under the GTR+I+G nucleotide substitution model as recommended by the jModeltest program [37], the heuristic tree search was performed using the SPR branch-swapping algorithm and branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test [38].

Analysis of spatiotemporal dispersion pattern and demographic history

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{MRCA}, years), the ancestral geographic movements, and the mode and rate (r, years-1) of population growth of HIV-1 subtype G clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 [39,40] with BEAGLE to improve run-time [41]. Analyses were performed under a GTR+I+G nucleotide substitution model. The temporal scale of the evolutionary process was estimated from the sampling dates of the sequences using a relaxed uncorrelated lognormal

molecular clock model and a uniform prior on clock rate ($1.5–3.0 \times 10^{-3}$ subst/site/year) [42]. Migration events throughout the phylogenetic history were inferred using a reversible discrete Bayesian phylogeographic model [43], in which all possible reversible exchange rates between locations were equally likely, and a CTMC rate reference prior [44]. Changes in effective population size through time were initially estimated using a flexible Bayesian Skyline coalescent model [45] and estimates of the population growth rate were subsequently obtained using the parametric model (logistic, exponential or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and stepping-stone sampling (SS) methods [46]. MCMC chains were run for $10–100 \times 10^6$ generations. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, using the TRACER v1.6 program [47]. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.4.0 [48].

Results

Identification of major HIV-1 subtype G clades

The likelihood mapping analysis and the transitions/transversions versus divergence plots indicates that all datasets used in our study retained enough phylogenetic signal for consistent phylogenetic inferences and no evidence of substitution saturation (S1 Fig). The ML phylogenetic tree of 578 HIV-1 subtype G *pol* sequences (566 classified as subtype G and 12 classified as CRF14_BG in the Los Alamos HIV Sequence Database) isolated in Portugal, Spain and 12 countries from West and Central Africa between 1992 and 2013 (Table 1) points to a clear phylogeographic subdivision of viral strains (Fig 1). Subtype G sequences from continental western African countries branched mostly in two large monophyletic clades (G_{WA-I} and G_{WA-II}) that were nested among the most basal clades from Central and West Central Africa (G_{CA}), consistent with our previous findings [21]. Although some subtype G sequences from Cape Verde ($n = 10$) also branched within the G_{WA-I} clade; most sequences from this insular West African country ($n = 48$) branched together with most subtype G sequences from Portugal ($n = 102$) in a distinct monophyletic clade (G_{CV-PT}) nested among basal G_{CA} lineages. All CRF14_BG sequences and several subtype G *pol* sequences from Portugal ($n = 78$) and Cape Verde ($n = 7$) formed a highly supported sub-cluster (CRF14_BG-like) within the G_{CV-PT} clade. According to the relative prevalence of the distinct subtype G clades, we can describe four basic molecular epidemiologic scenarios (Fig 2 and S2 Table): 1) basal G_{CA} clades are the predominant subtype G lineages circulating in countries from Central (90%) and West-Central (50%) African regions; 2) the G_{WA-I} clade was the predominant lineage detected in Nigeria (78%), Senegal (50%) and Benin (47%); 3) the G_{WA-II} clade predominates in Togo/Ghana (84%); and 4) the G_{CV-PT} clade was the dominant lineage in Cape Verde (80%) and Portugal (95%).

Spatiotemporal dispersal pattern of the HIV-1 G_{CV-PT} clade

To reconstruct the subtype G migrations between Africa and Portugal, all subtype G sequences belonging to the G_{CV-PT} clade (excluding the CRF14_BG-like sub-clade) ($n = 69$) were combined with basal G_{CA} strains of Central African origin ($n = 73$). Sequences were divided in six geographical locations, as those neighboring countries from Central and West-Central Africa comprising few samples ($n < 20$) were grouped into the same location state (S3 Table), and subjected to Bayesian phylogeographic analysis According to the Bayesian MCMC analysis, the most probable root location of the subtype G clade was placed in Central Africa (posterior state
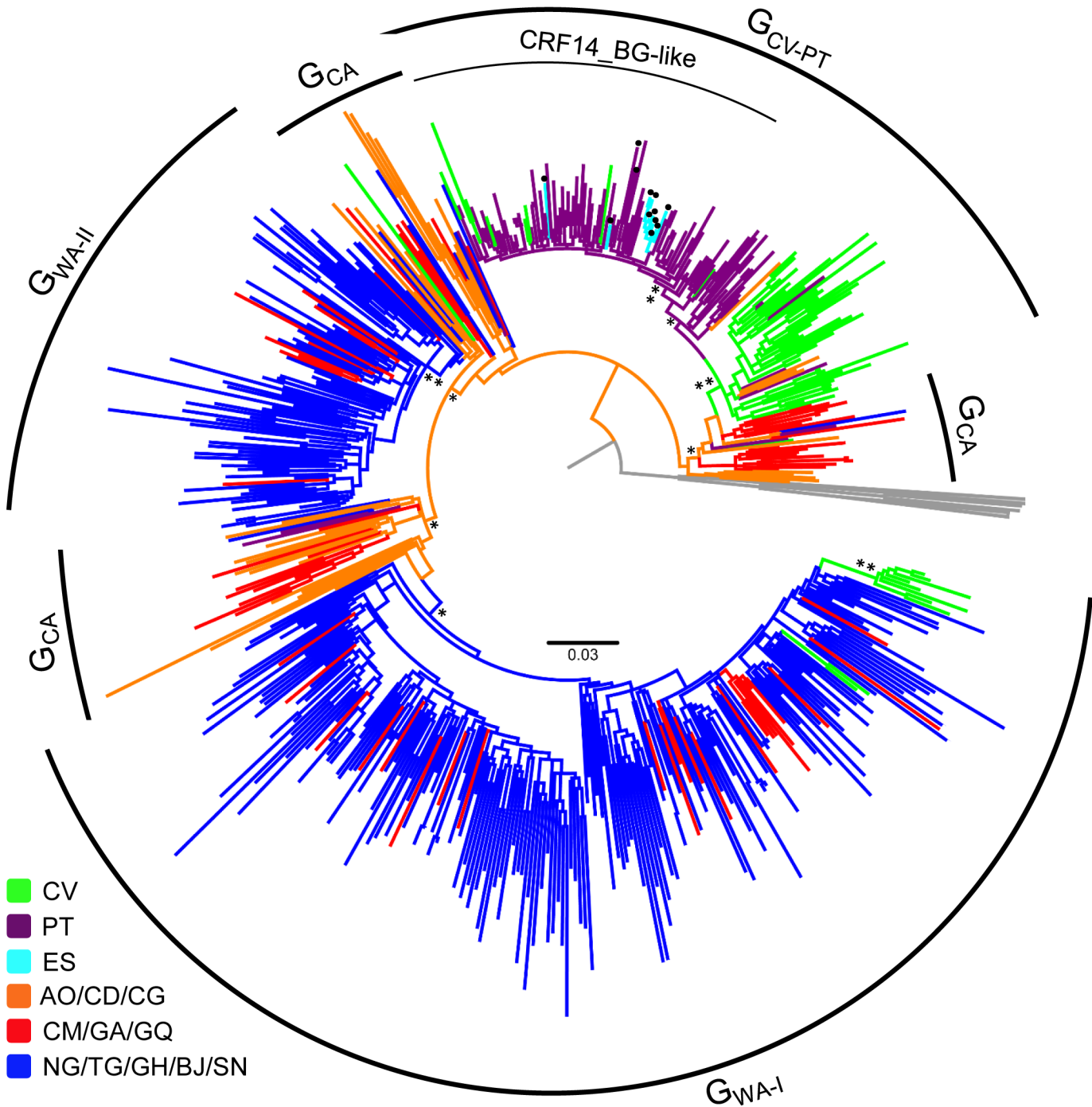
Fig 1. ML tree of the HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from Central Africa, West Africa, Portugal and Spain. Branches are colored according to the geographic origin of each sequence as indicated in the legend (bottom left). Arcs indicate the positions of major subtype G clades circulating in Central Africa (G_{CA}), West Africa (G_{WA-I} and G_{WA-II}) and Cape Verde/Portugal (G_{CV-PT}) and the position of the subclade that comprises all CRF14_BG reference sequences (CRF14_BG-like). Black dots indicate the positions of the reference sequences classified as CRF14_BG based on full-length genome analysis. Asterisks point to key nodes with relatively high (*, *a*LRT > 0.80) and high (**, *a*LRT > 0.90) support. The tree was rooted using HIV-1 subtype A-D reference sequences. The branch lengths are drawn to scale with the bar at the center indicating nucleotide substitutions per site. AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo; CM/GA/GQ: Cameroon/Gabon/Equatorial Guinea; NG/TG/GH/BJ/SN: Nigeria/Togo/Ghana/Benin/Senegal; CV: Cape Verde; PT: Portugal; ES: Spain.

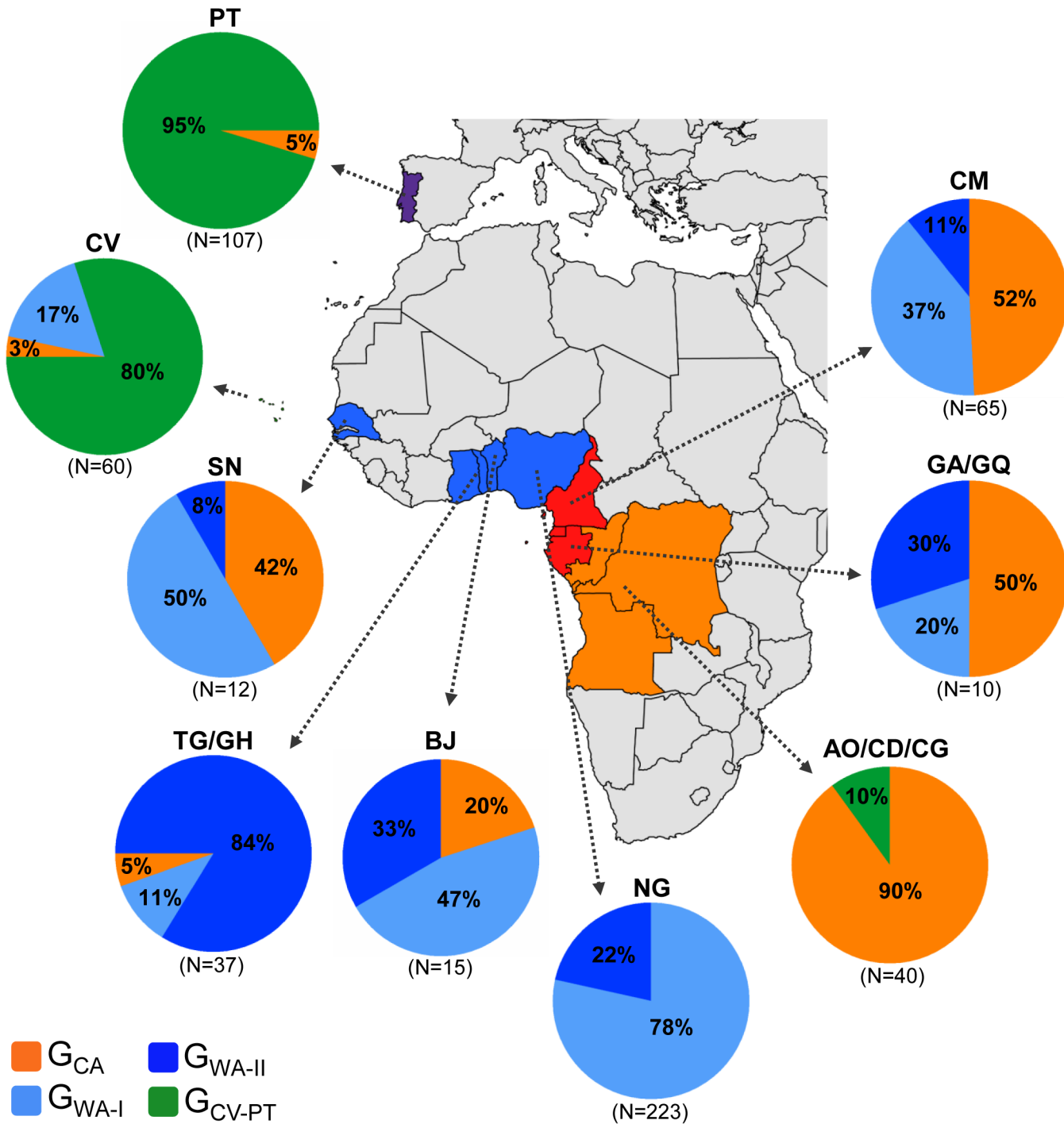doi:10.1371/journal.pone.0127384.g001

Fig 2. Prevalence of G_{CV-PT}, G_{WA-I}, G_{WA-II} and G_{CA} clades among HIV-1 subtype G infected individuals from different countries, estimated from phylogenetic analyses presented in Fig 1. The total number of subtype G sequences analyzed in each locality is indicated. Each clade is represented by a color as indicated at the legend. AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo; BJ: Benin; CM: Cameroon; CV: Cape Verde; GA/GQ: Gabon/Equatorial Guinea; NG: Nigeria; PT: Portugal; TG/GH: Togo/Ghana; SN: Senegal.

doi:10.1371/journal.pone.0127384.g002

probability, *PSP* = 1), and the onset date of this clade was estimated to be 1964 (95% HPD: 1937–1978) (Fig 3). Sequences from Cape Verde branched at the base of the G_{CV-PT} clade,

Fig 3. Time-scaled Bayesian MCC tree of HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from the G_{CA} and G_{CV-PT} clades. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). The arc indicates the position of the G_{CV-PT} clade. Key nodes corresponding to the MRCA of the Cape Verdean and Portuguese G_{CV-PT} lineages are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each lineage is indicated at the bottom left. Asterisks point to key nodes with relatively high (*, *PP* > 0.80) and high (**, *PP* > 0.90) posterior probability support. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a

whereas most sequences from Portugal branched in a monophyletic cluster nested within the Cape Verdean sequences ([Fig 3](#)). This analysis suggests that the G_{CV-PT} clade most probably migrates from Central Africa to Cape Verde (*PSP* = 0.68) at 1977 (95% HPD: 1972–1982) and rapidly moved from Cape Verde to Portugal in 1979 (95% HPD: 1974–1984). A few additional exchanges of the G_{CV-PT} clade between Cape Verde and Portugal were detected at later times ([Fig 3](#)).

Demographic history of the HIV-1 G_{CV-PT} clade

To reconstruct the demographic history of the G_{CV-PT} clade, all subtype G sequences from Cape Verde (*n* = 41) and Portugal (*n* = 24) that branched within this clade (excluding the CRF14_BG-like sub-clade) were selected. In agreement with our previous analysis, most sub-type G sequences from Portugal branched in a sub-cluster nested among basal Cape Verdean sequences ([Fig 4A](#)). This new analysis, however, supports a relatively more recent time-scale than previous estimations. According to this new analysis, the G_{CV-PT} clade probably arose in Cape Verde (*PSP* = 0.76) in 1984 (95% HPD: 1979–1989) and was rapidly disseminated to Portugal in 1987 (95% HPD: 1983–1990). Two additional migrations of the G_{CV-PT} clade from Cape Verde to Portugal and one migration event from Portugal to Cape Verde were also detected, in agreement with our previous analysis ([Fig 4A](#)). The Bayesian skyline plot (BSP) analysis suggests that the G_{CV-PT} clade experienced a fast exponential growth during the 1980s and 1990s, followed by a more recent stabilization since the early 2000s ([Fig 4B](#)). According to the logistic growth coalescent model, selected as the best-fit demographic model for the G_{CV-PT} clade (log BF > 10) ([S4 Table](#)), the mean growth rate of this subtype G clade was 0.52 year^{-1} (95% HPD: 0.32–0.77 year^{-1}) ([Fig 4C](#)).

Origin of the CRF14_BG clade

To investigate the origin of the parental subtype lineage that gave rise to the CRF14_BG, all *pol* sequences that branched within the CRF14_BG-like subclade (*n* = 97) were combined with sequences from clades G_{CA} and G_{CV-PT}. The overall topology and temporal structure of the Bayesian MCC trees remains conserved after inclusion of the CRF14_BG-like subclade, but placed most of the posterior root state probability mass of the G_{CV-PT} clade in Portugal (*PSP* = 0.55–0.81) ([Fig 5](#) and [S5 Table](#)). Both Bayesian MCMC analyses showed that all CRF14_BG-like sequences formed a well-supported sub-cluster (*PP* > 0.90) nested among basal subtype G Portuguese sequences within the G_{CV-PT} radiation ([Fig 5](#)). Those analyses support that the CRF14_BG-like clade most probably arose in Portugal (*PSP* = 1) and was later disseminated at multiple times from Portugal to both Spain and Cape Verde ([Fig 5](#)). The T_{MRCA} of the CRF14_BG-like clade was traced to 1986 (95% HPD: 1982–1991) when basal G_{CA} strains were included in the analysis ([Fig 5A](#)), and to 1991 (95% HPD: 1988–1994) when basal G_{CA} strains were not included ([Fig 5B](#)).

Discussion

This and our previous study [21] indicate that the HIV-1 subtype G likely originated in Central Africa around the middle-late 1960s and began to be disseminated to Western and West-Central Africa from the middle 1970s onwards. Some of the subtype G strains disseminated out of Central Africa fueled secondary outbreaks that led to the origin of regional-specific subtype G
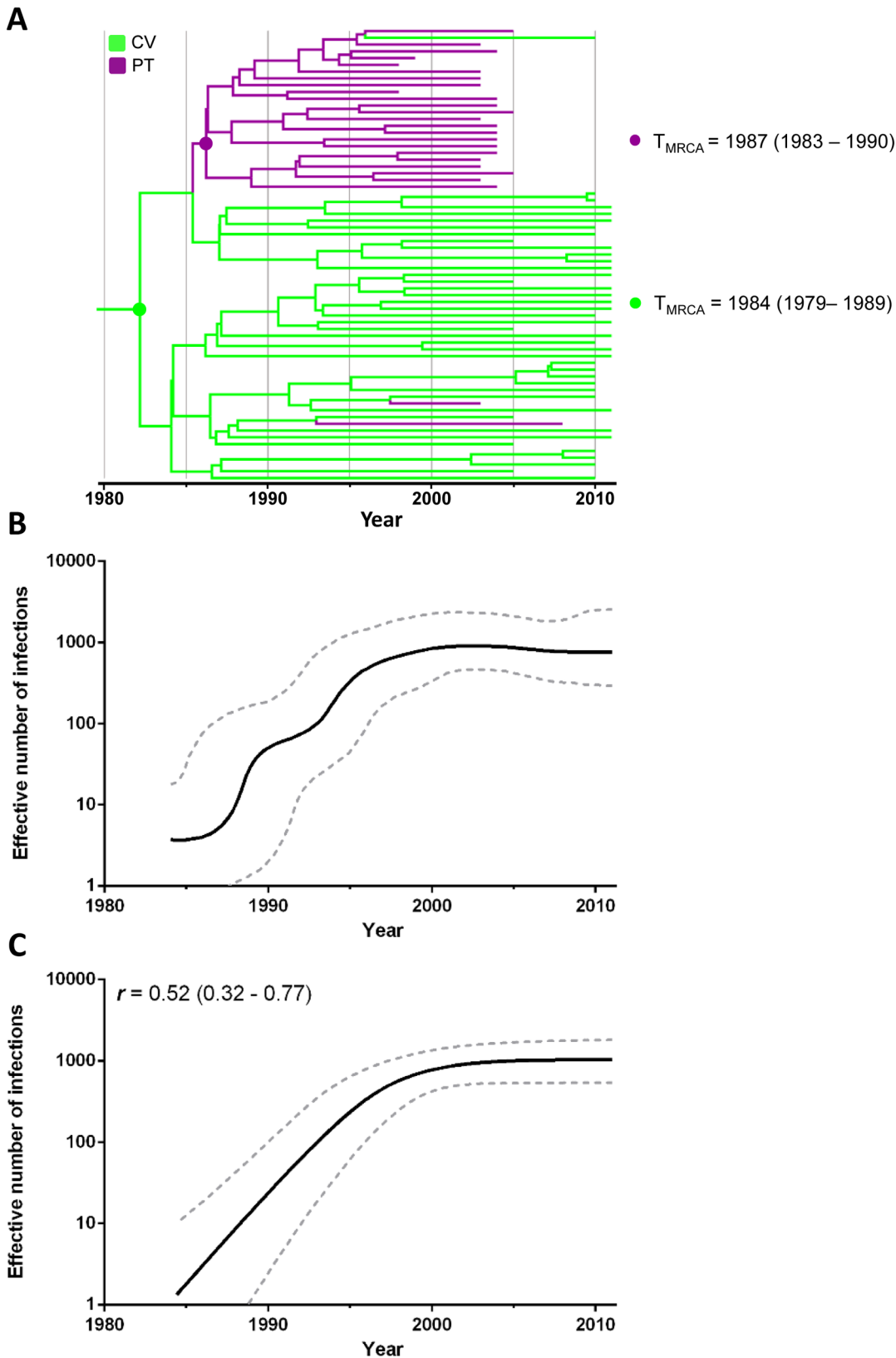
Fig 4. Demographic history of the HIV-1 G$_{CV-PT}$ clade circulating in Cape Verde and Portugal. A) Time-scaled Bayesian MCC tree of the HIV-1 G$_{CV-PT}$ clade. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). Key nodes

corresponding to the MRCA of the Cape Verde and Portuguese G_{CV-PT} lineages are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each lineage is indicated at right. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a relaxed molecular clock. B and C) Effective number of infections (y-axis; log10 scale) through time (x-axis; calendar years) estimated using Bayesian skyline (B) and logistic growth (C) coalescent models. Median (solid line) and 95% HPD intervals (dashed lines) of the effective number of infections estimated through time are shown in each graphic. The median growth rate (with the corresponding 95% HPD interval) of G_{CV-PT} clade estimated under the logistic growth model is indicated in the upper left corner.

clades. The major subtype G clades detected in our previous study in West Africa were the G_{WA-I} (that most probably emerged in Nigeria around the middle 1970s) and the G_{WA-II} (that most probably emerged in Togo or Ghana around the late 1970s) [21]. In the present study we identified a novel major clade (G_{CV-PT}) that probably emerged between the late 1970s and the middle 1980s and circulates in Cape Verde and Portugal.

The G_{CV-PT} clade comprises 95% and 80% of HIV-1 subtype G *pol* sequences from Portugal and Cape Verde included in our study, respectively. Within the G_{CV-PT} radiation, most sequences from Portugal (73%) branched in a monophyletic subclade together with the CRF14_BG reference sequences, whereas the remaining Portuguese sequences branched at the base of the CRF14_BG-like subclade. This clearly indicates that the G_{CV-PT} clade is the parental subtype G lineage of the CRF14_BG variant and that the CRF14_BG clade is probably more prevalent in Portugal than the parental G_{CV-PT} clade, consistent with previous findings [25]. It is also important to note that a small fraction of G_{CV-PT} *pol* sequences from Cape Verde (15%) also branched within the CRF14_BG-like clade, indicating that this recombinant lineage not only circulates in Portugal and Spain, but also in Cape Verde. Full-length genome analyses of Cape Verdean HIV-1 subtype G *pol* sequences that branched within the CRF14_BG-like subclade should be performed to confirm this hypothesis.

The phylogeographic analyses that combined subtype G sequences of the G_{CV-PT} clade (with exception of the CRF14_BG-like lineage) and basal G_{CA} clades consistently pointed to Cape Verde as the most probable root location of the G_{CV-PT} clade (*PSP* = 0.68–0.76). When CRF14_BG-like sequences are included, the root location of the G_{CV-PT} clade was most probably placed in Portugal (*PSP* = 0.55–0.81). It has been shown that convenience sampling (particularly sampling heterogeneity) can obfuscate the accurate estimation of ancestral spatial locations based on standard phylogeographic continuous-time Markov chain implementation [49]. When CRF14_BG-like sequences are included, the number of Portuguese sequences (n = 104) far exceeds the number of Cape Verdean sequences (n = 48) within the G_{CV-PT} clade and such a larger sample from Portugal may results in the higher support for this location as the origin of that clade. Thus, according to the more balanced data sets the founder G_{CV-PT} ancestor probably moved from Central Africa to Cape Verde and later passed from Cape Verde to Portugal.

Whereas the inclusion of the CRF14_BG-like sequences has a great impact on estimation of the G_{CV-PT} ancestral root location, ancestral root ages were mainly influenced by the inclusion of basal G_{CA} clades. The median T_{MRCA} of the G_{CV-PT} clade was traced to the late 1970s when basal G_{CA} clades were included, and to the middle 1980s when those basal sequences were not included (S5 Table). Similarly, the T_{MRCA} of the CRF14_BG clade moved from the middle 1980s to the early 1990s when G_{CA} clades were removed from the analysis (S5 Table). This suggests that inclusion of basal lineages from Central Africa tend to produce slightly older internal node ages, although no significant changes are observed in the mean estimated substitution rates (S5 Table). This observation, however, should be interpreted with caution because those T_{MRCA} estimates displayed a considerable overlap of the confidence interval and thus should not be regarded as statistically different.
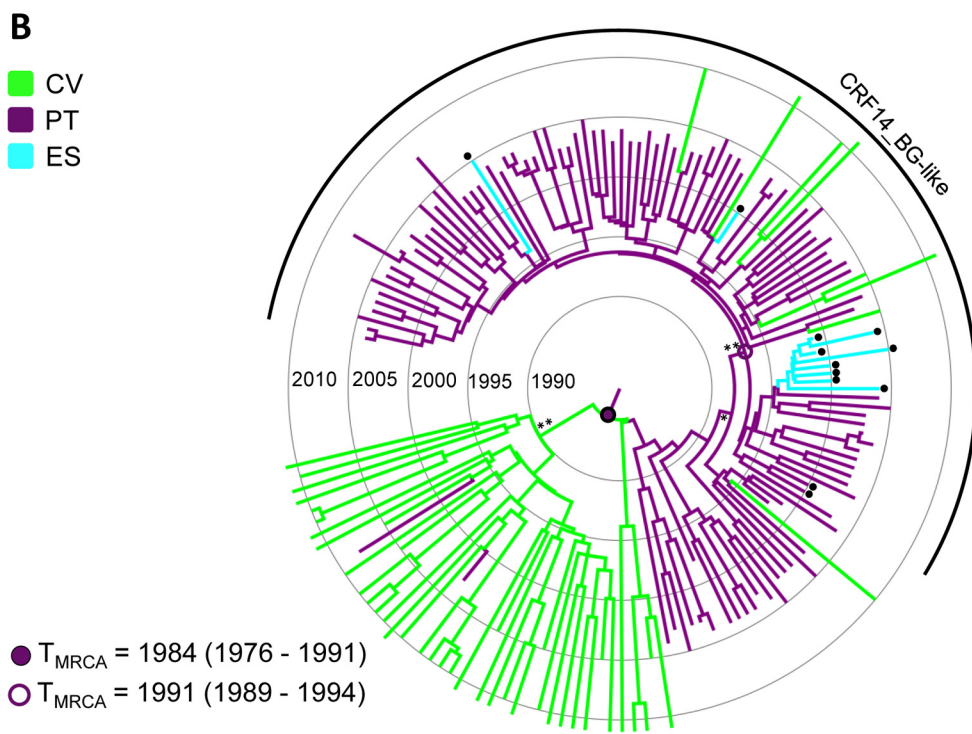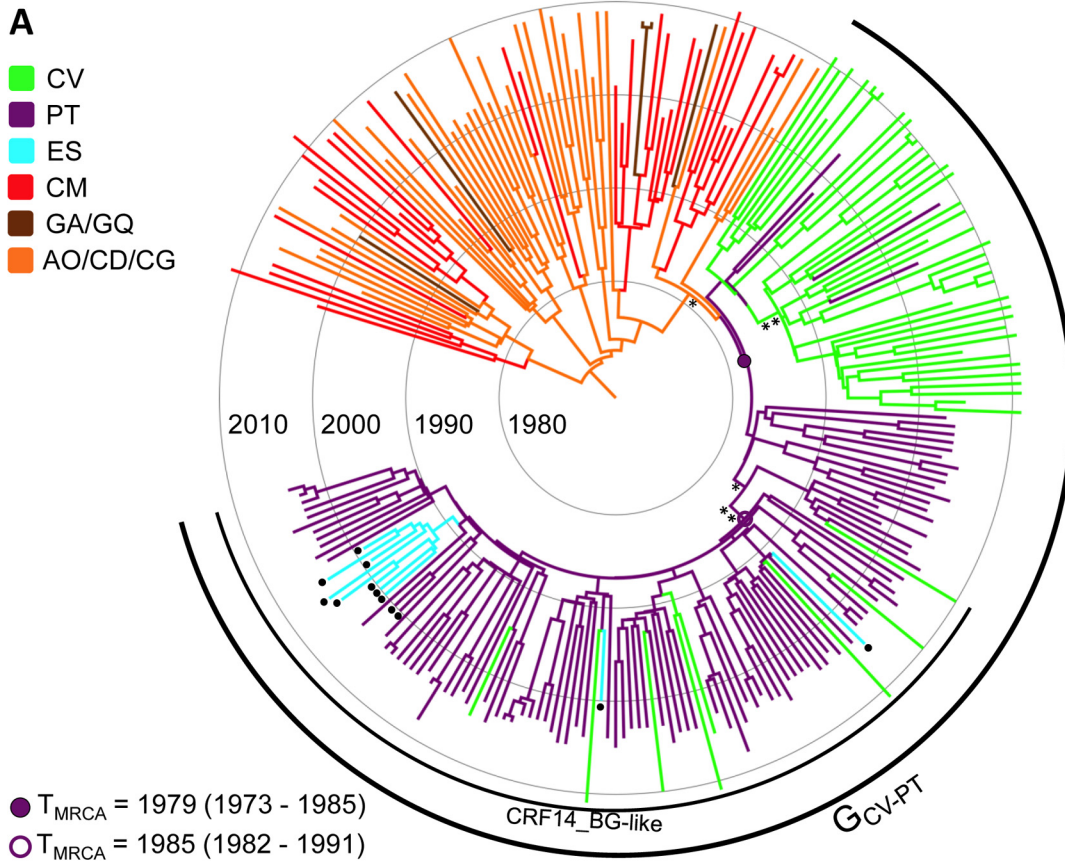
Fig 5. Time-scaled Bayesian MCC tree of HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from the G_{CA}, G_{CV-PT} and CRF14_BG-like clades.
Sequences that branched within the CRF14_BG-like subclade were combined with sequences from G_{CA} and G_{CV-PT} clades (A) or only G_{CV-PT} clade (B). Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). Arcs indicate the positions of G_{CV-PT} and CRF14_BG-like clades. Nodes corresponding to the MRCA of those clades are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each clade is indicated at the bottom left. Black dots indicate the position of the CRF14_BG reference sequences. Asterisks point to key nodes with high relatively high (*, *PP* > 0.80) and high (**, *PP* > 0.90) posterior probability support. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a relaxed molecular clock. AO/CD/CG: Angola/ Democratic Republic of Congo /Republic of Congo; CM: Cameroon; CV: Cape Verde; GA/GQ: Gabon/Equatorial Guinea; PT: Portugal; ES: Spain.

Regardless the precise root age, our phylogeographic analyses support a nearly simultaneous introduction and concurrent dissemination of the G_{CV-PT} clade in Cape Verde and Portugal. Our phylogeographic analyses based on balanced datasets suggest that the G_{CV-PT} clade started to be disseminated in Portugal only a couple of years later than the estimated introduction of the virus into Cape Verde. Of note, the estimated time-frame (1977–1984) for introduction and dissemination of the G_{CV-PT} clade in Cape Verde and Portugal was preceded by a phase of negative migratory outflow in Angola [50], associated to the exodus of thousands of Portuguese citizens of European and African ethnicity from Angola after the country independence in 1974. This may have fueled the chance exportation of the G_{CV-PT} ancestor strain from Angola into Cape Verde and its rapid dissemination to Portugal, thus suggesting that the global route of spread of the G_{CV-PT} clade was probably laid out along the colonial history ties, as has been previously demonstrated for the HIV-2 group A [49].

Despite the continuous and extensive migration of people between Angola, Cape Verde, and Portugal [28,51], subtype G strains sampled in those Portuguese-speaking countries retain a high phylogeographic structure with relative few viral exchanges among them. We have detected a total of: 1) four independent introductions of G_{CA} strains from Central Africa into Portugal, 2) three introductions of G_{CA} strains from Central Africa into Cape Verde, 3) three introductions of G_{CV-PT} strains from Cape Verde into Portugal, and 4) one G_{CV-PT} migration and five CRF14_BG introductions from Portugal into Cape Verde. Although the continuous viral exchanges among these countries may suppose a risk to the emergence of new country-specific subtype G lineages, most viral introductions seem to have failed to sustain new local subtype G epidemics with exception of the G_{CV-PT} founder strain.

According to our analysis, the G_{CV-PT} clade displayed a logistic population growth pattern characterized by an initial phase of exponential growth with a median rate of 0.52 year^{-1} (95% HPD: 0.32–0.77 year^{-1}), followed by a decline in growth rate since the early 2000s. The median estimated logistic growth rate of the G_{CV-PT} clade was similar to that estimated for basal G_{CA} clades in Central Africa (0.47 year^{-1}) [21] and the G_{CU} clade circulating in Cuba (0.55 year^{-1}) [52]; but lower than those previously estimated for the G_{WA-I} (0.75 year^{-1}) and G_{WA-II} (0.95 year^{-1}) clades circulating in continental West African countries [21] (S6 Table). The differential growth rates detected among different subtype G clades could be associated to clade-specific or ecological-specific differences in viral transmissibility. Further studies should be performed to understand whether the G_{WA-I} and G_{WA-II} clades introduced into continental West Africa displayed a higher intrinsic transmissibility or encountered more favorable epidemiological conditions for local and regional expansion than those disseminated within Central Africa, Cape Verde, Cuba and Portugal.

A previous study concluded that the CRF14_BG emerged in Portugal in the early 1990s and then spread to the North of Spain in late 1990s following the mobility of HIV-infected IDUs [27]. Our phylogeographic analyses indicate that the CRF14_BG clade probably arose in Portugal between the middle 1980s and the early 1990s, which is fully consistent with the previous estimation and with epidemiological data showing that CRF14_BG was already circulating in Lisbon in 1993 [27]. According to this estimate, the recombinant ancestor of the CRF14_BG

clade was generated about five years after the estimated arrival of the parental G_{CV-PT} clade into Portugal, thus indicating a very rapid generation of BG recombinants in this country. After a period of local dissemination within Portugal, the CRF14_BG clade was dispersed not only from Portugal to Spain, but also probably to Cape Verde at multiple times.

In summary, this study reveals that most HIV-1 subtype G infections in Cape Verde and Portugal have resulted from the local dissemination of a single clade (here called G_{CV-PT}) that probably emerged after a single migration of the virus out of Central Africa into Cape Verde between the late 1970s and the middle 1980s. Dispersion of the G_{CV-PT} clade seems to have been shaped by the historical and ongoing human population movements between Angola, Cape Verde and Portugal,. Our data also highlight that once introduced in Portugal, the G_{CV-PT} was disseminated in the local population and probably recombined with local preexisting subtype B variants, originating the CRF14_BG clade. These findings offer important insights to understanding the origin and current characteristics of the HIV-1 subtype G and CFR14_BG epidemics in Cape Verde and Portugal.

Supporting Information

S1 Fig. Analyses of phylogenetic signal and substitution saturation. (A—E) Likelihood maps of 10,000 random quartets made from every HIV-1 subtype G dataset used in this study as indicated in the figure. The triangles display the distribution (left) and percentage (right) of dots representing the likelihoods of the three possible tree topologies for a group of four sequences (quartets) randomly selected from the dataset. The tree-like, star-like and network-like phylogenetic signals are represented by the dots localized on the vertices, center and on the laterals, respectively. Fully resolved (tree-like) tree topologies ranged from 0.77 (CRF14-like-CV-PT) to 0.93 (G-CA + G-WA + G-CV-PT), thus indicating enough phylogenetic signal for consistent phylogenetic inferences in all datasets. (F—J) Substitution saturation plots of the datasets used in this study as depicted in the figure. The ordinate corresponds to the observed proportion of transitions (s, green) and transversions (v, blue) while the abscissa refers to the distance calculated using the GTR substitution model. The central lines of each plot correspond to the quadratic nonlinear regressions of the data. CA—Central Africa, WA—West Africa, CV—Cape Verde, PT—Portugal. All analyses indicated an absence of substitution saturation in the data set explored since the plots did not reach an evident plateau nor the transversions outnumbered transitions.
(PDF)

S1 Table. GenBank accession numbers of HIV-1 subtype G *pol* sequences described in Table 1.
(PDF)

S2 Table. Distribution of HIV-1 subtype G *pol* sequences across major regional clades circulating in Central/West-Central Africa (G_{CA}), West Africa (G_{WA-I} and G_{WA-II}) and Cape Verde/Portugal (G_{CV-PT}). AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo. GA/GQ: Gabon/Equatorial Guinea. GH/TG: Ghana/Togo. PT/ES: Portugal/Spain.
(PDF)

S3 Table. HIV-1 subtype G *pol* dataset used for Bayesian phylogeographic analysis. [a]The number of subtype G *pol* fragments recovered from full-length HIV-1 CRF14_BG reference sequences is indicated in parenthesis. DRC: Democratic Republic of Congo.
(PDF)

S4 Table. Best fit demographic model for HIV-1 G$_{CV-PT}$ clade. Log marginal likelihood (ML) estimates for the logistic (Log), exponential (Expo) and expansion (Expa) growth demographic models obtained using the path sampling (PS) and stepping-stone sampling (SS) methods. The Log Bayes factor (BF) is the difference of the Log ML between of alternative (H1) and null (H0) models (H1/H0). Log BFs > 3 indicates that model H1 is more strongly supported by the data than model H0.
(PDF)

S5 Table. Bayesian estimates of the age and root location of the most recent common ancestor (MRCA) of major HIV-1 subtype G (G$_{CV-PT}$) and BG (CRF14_BG) clades circulating in Cape Verde and Portugal. [a] substitutions/site/year. CV: Cape Verde. PT: Portugal. TMRCA: time of the most recent common ancestor. PSP: posterior state probability.
(PDF)

S6 Table. Evolutionary and demographic parameters estimated for major HIV-1 subtype G clades circulating in Central/West-Central Africa (G$_{CA}$), West Africa (G$_{WA-I}$ and G$_{WA-II}$), Cuba (G$_{CU}$) and Cape Verde/Portugal (G$_{CV-PT}$). [a] Data from Delatorre et al [21]. [b] Data from Delatorre et al [52]. [c] Estimated at this study.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: GB MGM ED IIMPA. Performed the experiments: IIMPA ED GB. Analyzed the data: GB ED IIMPA MGM MLG. Contributed reagents/materials/analysis tools: IIMPA ED MLG MGM GB. Wrote the paper: GB ED IIMPA MGM MLG.

References

1. Archer J, Robertson DL (2007) Understanding the diversification of HIV-1 groups M and O. Aids 21: 1693–1700. PMID: 17690566

2. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids 25: 679–689. doi: 10.1097/QAD.0b013e328342ff93 PMID: 21297424

3. Oliveira V, Bartolo I, Borrego P, Rocha C, Valadas E, Barreto J, et al. (2012) Genetic diversity and drug resistance profiles in HIV type 1- and HIV type 2-infected patients from Cape Verde Islands. AIDS Res Hum Retroviruses 28: 510–522. doi: 10.1089/AID.2011.0174 PMID: 21902592

4. de Pina-Araujo I, Guimaraes ML, Bello G, Vicente AC, Morgado MG (2014) Profile of the HIV epidemic in Cape Verde: molecular epidemiology and drug resistance mutations among HIV-1 and HIV-2 infected patients from distinct islands of the archipelago. PLOS One 9: e96201. doi: 10.1371/journal.pone.0096201 PMID: 24763617

5. Peeters M, Esu-Williams E, Vergne L, Montavon C, Mulanga-Kabeya C, Harry T, et al. (2000) Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution. AIDS Res Hum Retroviruses 16: 315–325. PMID: 10716369

6. Agwale SM, Zeh C, Robbins KE, Odama L, Saekhou A, Edubio A, et al. (2002) Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials. Vaccine 20: 2131–2139. PMID: 11972982

7. Ojesina AI, Sankale JL, Odaibo G, Langevin S, Meloni ST, Sarr AD, et al. (2006) Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. AIDS Res Hum Retroviruses 22: 770–779. PMID: 16910833

8. Sankale JL, Langevin S, Odaibo G, Meloni ST, Ojesina AI, Olaleye D, et al. (2007) The complexity of circulating HIV type 1 strains in Oyo state, Nigeria. AIDS Res Hum Retroviruses 23: 1020–1025. PMID: 17725419

9. Chaplin B, Eisen G, Idoko J, Onwujekwe D, Idigbe E, Adewole I, et al. (2011) Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy. AIDS Res Hum Retroviruses 27: 71–80. doi: 10.1089/aid.2010.0050 PMID: 20964479

10. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, Conradie F, et al. (2011) HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. Lancet Infect Dis 11: 750–759. doi: 10.1016/S1473-3099(11)70149-9 PMID: 21802367

11. Ajoge HO, Gordon ML, Ibrahim S, Shittu OS, Ndung'u T, Olonitola SO (2012) Drug resistance pattern of HIV type 1 isolates sampled in 2007 from therapy-naive pregnant women in North-Central Nigeria. AIDS Res Hum Retroviruses 28: 115–118. doi: 10.1089/AID.2011.0115 PMID: 21568761

12. Imade GE, Sagay AS, Chaplin B, Chebu P, Musa J, Okpokwu J, et al. (2014) Short communication: Transmitted HIV drug resistance in antiretroviral-naive pregnant women in north central Nigeria. AIDS Res Hum Retroviruses 30: 127–133. doi: 10.1089/AID.2013.0074 PMID: 24164431

13. Bartolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, Fonseca M, et al. (2009) Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. Infect Genet Evol 9: 672–682. doi: 10.1016/j.meegid.2008.05.003 PMID: 18562253

14. Afonso JM, Bello G, Guimaraes ML, Sojka M, Morgado MG (2012) HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. PLOS ONE 7: e42996. doi: 10.1371/journal.pone.0042996 PMID: 22952625

15. Bartolo I, Zakovic S, Martin F, Palladino C, Carvalho P, Camacho R, et al. (2014) HIV-1 diversity, transmission dynamics and primary drug resistance in Angola. PLOS One 9: e113626. doi: 10.1371/journal.pone.0113626 PMID: 25479241

16. Chamberland A, Diabate S, Sylla M, Anagounou S, Geraldo N, Zannou DM, et al. (2012) Transmission of HIV-1 drug resistance in Benin could jeopardise future treatment options. Sex Transm Infect 88: 179–183. doi: 10.1136/sextrans-2011-050209 PMID: 22158948

17. Mamadou S, Montavon C, Ben A, Djibo A, Rabiou S, Mboup S, et al. (2002) Predominance of CRF02-AG and CRF06-cpx in Niger, West Africa. AIDS Res Hum Retroviruses 18: 723–726. PMID: 12167280

18. Charpentier C, Bellecave P, Cisse M, Mamadou S, Diakite M, Peytavin G, et al. (2011) High prevalence of antiretroviral drug resistance among HIV-1-untreated patients in Guinea-Conakry and in Niger. Antivir Ther 16: 429–433. doi: 10.3851/IMP1754 PMID: 21555827

19. Yaotse DA, Nicole V, Roch NF, Mireille PD, Eric D, Martine P (2009) Genetic characterization of HIV-1 strains in Togo reveals a high genetic complexity and genotypic drug-resistance mutations in ARV naive patients. Infect Genet Evol 9: 646–652. doi: 10.1016/j.meegid.2009.04.002 PMID: 19460333

20. Dagnra AY, Vidal N, Mensah A, Patassi A, Aho K, Salou M, et al. (2011) High prevalence of HIV-1 drug resistance among patients on first-line antiretroviral treatment in Lome, Togo. J Int AIDS Soc 14: 30. doi: 10.1186/1758-2652-14-30 PMID: 21663632

21. Delatorre E, Mir D, 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 PMID: 24918930

22. Esteves A, Parreira R, Piedade J, Venenno T, Franco M, Germano de Sousa J, et al. (2003) Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. AIDS Res Hum Retroviruses 19: 511–517. PMID: 12892060

23. Esteves A, Parreira R, Venenno T, Franco M, Piedade J, Germano de Sousa J, et al. (2002) Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. AIDS Res Hum Retroviruses 18: 313–325. PMID: 11897032

24. Palma AC, Araujo F, Duque V, Borges F, Paixao MT, Camacho R (2007) Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. Infect Genet Evol 7: 391–398. PMID: 17360244

25. Abecasis AB, Martins A, Costa I, Carvalho AP, Diogo I, Gomes P, et al. (2011) Molecular epidemiological analysis of paired pol/env sequences from Portuguese HIV type 1 patients. AIDS Res Hum Retroviruses 27: 803–805. doi: 10.1089/AID.2010.0312 PMID: 21198411

26. Delgado E, Thomson MM, Villahermosa ML, Sierra M, Ocampo A, Miralles C, et al. (2002) Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. J Acquir Immune Defic Syndr 29: 536–543. PMID: 11981372

27. Bartolo I, Abecasis AB, Borrego P, Barroso H, McCutchan F, Gomes P, et al. (2011) Origin and epidemiological history of HIV-1 CRF14_BG. PLOS One 6: e24130. doi: 10.1371/journal.pone.0024130 PMID: 21969855

28. Carling J (2004) Emigration, Return and Development in Cape Verde: The Impact of Closing Borders. Popul Space Place 10: 113–132.

29. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797–3800. PMID: 16076886

30. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, 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. PMID: 9847317

31. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A 94: 6815–6819. PMID: 9192648

32. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504. PMID: 11934758

33. Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, et al. (2009) Mobyle: a new full web bioinformatics framework. Bioinformatics 25: 3005–3011. doi: 10.1093/bioinformatics/btp493 PMID: 19689959

34. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371–373. PMID: 11535656

35. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321. doi: 10.1093/sysbio/syq010 PMID: 20525638

36. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557–559. PMID: 15980534

37. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256. doi: 10.1093/molbev/msn083 PMID: 18397919

38. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552. PMID: 16785212

39. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320. PMID: 12136032

40. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214. PMID: 17996036

41. Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. Bioinformatics 25: 1370–1376. doi: 10.1093/bioinformatics/btp244 PMID: 19369496

42. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4: e88. PMID: 16683862

43. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLOS Comput Biol 5: e1000520. doi: 10.1371/journal.pcbi.1000520 PMID: 19779555

44. Ferreira MAR, M.A. S (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics 26: 355–368.

45. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192. PMID: 15703244

46. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013. PMID: 11371589

47. Rambaut A, Drummond A (2007) Tracer v1.6. Available from http://treebioedacuk/software/tracer/.

48. Rambaut A (2009) FigTree v1.4: Tree Figure Drawing Tool. Available from http://treebioedacuk/software/figtree/.

49. Faria NR, Hodges-Mameletzis I, Silva JC, Rodes B, Erasmus S, Paolucci S, et al. (2012) Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol 93: 889–899. doi: 10.1099/vir.0.038638-0 PMID: 22190015

50. Bello G, Afonso JM, Morgado MG (2012) Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. Infect Genet Evol 12: 1079–1086. doi: 10.1016/j.meegid.2012.03.014 PMID: 22484759

51. Dias P, Machado R, Bento AR, (SEF) SdEeF (2014) Relatório de Imigração, Fronteiras e Asilo—2013 (in Portuguese). Available at: http://sefstatsefpt/Docs/Rifa_2013pdf.

52. Delatorre E, Bello G (2013) Phylodynamics of the HIV-1 epidemic in Cuba. PLOS ONE 8: e72448. doi: 10.1371/journal.pone.0072448 PMID: 24039765