

Phylodynamics of the HIV-1 Epidemic in Cuba

Edson Delatorre, Gonzalo Bello*

Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

Abstract

Previous studies have shown that the HIV-1 epidemic in Cuba displayed a complex molecular epidemiologic profile with circulation of several subtypes and circulating recombinant forms (CRF); but the evolutionary and population history of those viral variants remains unknown. HIV-1 *pol* sequences of the most prevalent Cuban lineages (subtypes B, C and G, CRF18_cpx, CRF19_cpx, and CRFs20/23/24_BG) isolated between 1999 and 2011 were analyzed. Maximum-likelihood analyses revealed multiple introductions of subtype B ($n \geq 66$), subtype C ($n \geq 10$), subtype G ($n \geq 8$) and CRF18_cpx ($n \geq 2$) viruses in Cuba. The bulk of HIV-1 infections in this country, however, was caused by dissemination of a few founder strains probably introduced from North America/Europe (clades B_{CU-I} and B_{CU-II}), east Africa (clade C_{CU-I}) and central Africa (clades G_{CU}, CRF18_{CU} and CRF19_{CU}), or locally generated (clades CRFs20/23/24_BG). Bayesian-coalescent analyses show that the major HIV-1 founder strains were introduced into Cuba during 1985–1995; whereas the CRFs_BG strains emerged in the second half of the 1990s. Most HIV-1 Cuban clades appear to have experienced an initial period of fast exponential spread during the 1990s and early 2000s, followed by a more recent decline in growth rate. The median initial growth rate of HIV-1 Cuban clades ranged from 0.4 year⁻¹ to 1.6 year⁻¹. Thus, the HIV-1 epidemic in Cuba has been a result of the successful introduction of a few viral strains that began to circulate at a rather late time of the AIDS pandemic, but then were rapidly disseminated through local transmission networks.

Citation: Delatorre E, Bello G (2013) Phylodynamics of the HIV-1 Epidemic in Cuba. PLoS ONE 8(9): e72448. doi:10.1371/journal.pone.0072448

Editor: Chiyu Zhang, Institut Pasteur of Shanghai, Chinese Academy of Sciences, China

Received: June 14, 2013; **Accepted:** July 17, 2013; **Published:** September 9, 2013

Copyright: © 2013 Delatorre, Bello. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Public Health Service grants E-26/111.758/2012 from the FAPERJ and 472896/2012-1 from the CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gbellobr@gmail.com

Introduction

The global dissemination of the Human immunodeficiency virus type 1 (HIV-1) group M clade during the second half of the twentieth century has resulted in the generation of a diverse collection of genetic variants classified into subtypes, sub-subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs). The HIV-1 epidemic in the Americas is typically dominated by subtype B clade, although substantial proportions ($\geq 20\%$) of non-B subtype genetic forms are observed in Argentina, Brazil, Cuba and Uruguay [1].

Cuba displayed a unique HIV-1 molecular epidemiologic profile in the Americas characterized by the co-circulation of several subtypes (A1, B, C, F1, G, H and J), CRFs and URFs. Subtype B is the most prevalent variant ($\sim 33\text{--}40\%$), followed by CRF19_cpx ($\sim 20\text{--}28\%$), CRFs20/23/24_BG ($\sim 12\text{--}20\%$) CRF18_cpx ($\sim 7\text{--}10\%$), subtype C ($\sim 3\text{--}10\%$), and subtype G ($\sim 2\text{--}7\%$) [2,3,4,5,6,7]. It has been proposed that the presence of numerous Cuban military and civilian personnel in several sub-Saharan African countries, and particularly those stationed in Angola and neighboring countries between 1975 and 1991, have contributed to the introduction of multiple non-B HIV-1 subtypes into Cuba [2]. Some HIV-1 recombinants including CRF18_cpx and CRF19_cpx were probably also imported into Cuba directly from central Africa, since the parental viruses of these complex genetic forms were only detected in that African region [8,9]. Indeed, a few cases of CRF18_cpx and CRF19_cpx like viruses have been confirmed in Angola [10,11], Democratic Republic of Congo (DRC) [12,13], Republic of Congo [14,15], Central African Republic [16], and Cameroon [17,18,19]. Other HIV-1

recombinants including all three CRFs_BG, however, were probably generated locally by recombination between subtypes B and G already circulating in Cuba [20].

According to this model, most non-B subtype HIV-1 variants circulating in Cuba were probably introduced or locally generated after 1975. Despite the extensive knowledge about the molecular epidemiology of HIV-1 variants, the time-scale and epidemic history of most prevalent HIV-1 clades circulating in Cuba remains to be elucidated. In this study, we used a Bayesian coalescent-based method and a comprehensive data set of HIV-1 subtype B ($n = 322$), and non-B subtypes ($n = 420$) *pol* sequences of Cuban origin isolated between 1999 and 2011, to date the origin and reconstruct the demographic history of major HIV-1 variants circulating in Cuba.

Materials and Methods

HIV-1 Cuban sequence datasets

We downloaded all HIV-1 Cuban sequences covering the entire protease and partial reverse transcriptase (PR/RT) regions of the *pol* gene (nt 2253–3272 relative to HXB2 clone) classified as subtypes B ($n = 322$), C ($n = 49$), G ($n = 35$), CRF18_cpx ($n = 71$), CRF19_cpx ($n = 167$), and CRFs20/23/24_BG ($n = 118$) that were available at the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by March 2013. HIV-1 *pol* sequences were retrieved from both antiretroviral therapy naïve and HAART treated patients from different Cuban regions between 1999 and 2011, as described in previous studies [2,3,4,5,6]. Sequences were aligned using the CLUSTAL X program [21]. To avoid any bias on the

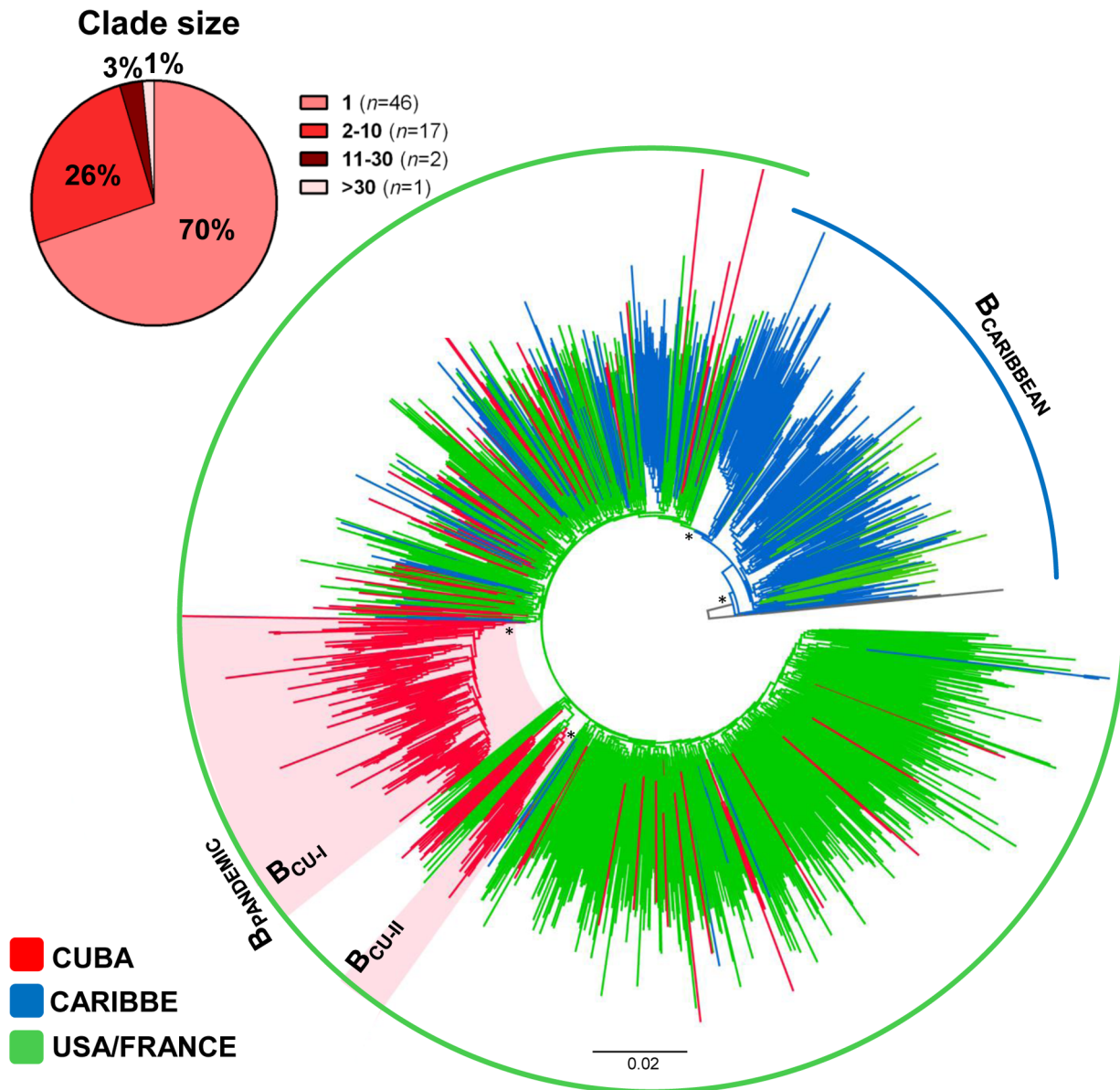


Figure 1. ML phylogenetic tree of HIV-1 subtype B *pol* (~1000 pb) sequences circulating in Cuba ($n=322$), US ($n=525$), France ($n=348$), and other Caribbean countries ($n=418$). The branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). The circular brackets highlight the position of the pandemic (B_{PANDEMIC} , green line) and non-pandemic ($B_{\text{CARIBBEAN}}$, blue line) HIV-1 subtype B clades. Shaded boxes highlight the position of the two major HIV-1 subtype B Cuban clades ($B_{\text{CU-I}}$ and $B_{\text{CU-II}}$). The number of Cuban sequences distributed accordingly to the clade size is shown (top left). Key nodes with αLRT support values >0.80 (*) and ≥ 0.90 (**) are indicated. The tree was rooted using HIV-1 subtype D reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g001

phylogenetic reconstructions, all sites with major antiretroviral drug resistance mutations in PR (30, 32, 46, 47, 48, 50, 54, 76, 82, 84, 88 and 90) or RT (41, 65, 67, 69, 70, 74, 100, 101, 103, 106, 115, 138, 151, 181, 184, 188, 190, 210, 215, 219 and 230) detected in at least two sequences were excluded from each alignment. All alignments are available from the authors upon request.

HIV-1 reference datasets

HIV-1 Cuban sequences were combined with reference sequences of diverse origin that matched the selected genomic region and were available at the Los Alamos HIV Sequence

Database. Subtype B Cuban sequences were aligned with reference sequences representative of the viral diversity in US ($n=525$), France ($n=348$) and the Caribbean ($n=417$) (Table S1). Subtype C Cuban sequences were aligned with representative sequences from central ($n=53$), eastern ($n=330$) and southern ($n=545$) African regions (Table S2). The HIV-1 subtype G Cuban sequences were combined with all available subtype G sequences of African origin ($n=437$) (Table S3). The CRF19_cpx Cuban sequences were aligned with all available CRF19_cpx sequences from other countries ($n=3$) and subtype D sequences of African origin ($n=1,112$) (Table S4). Finally, the HIV-1 CRF18_cpx and CRFs_BG Cuban sequences were combined with all available

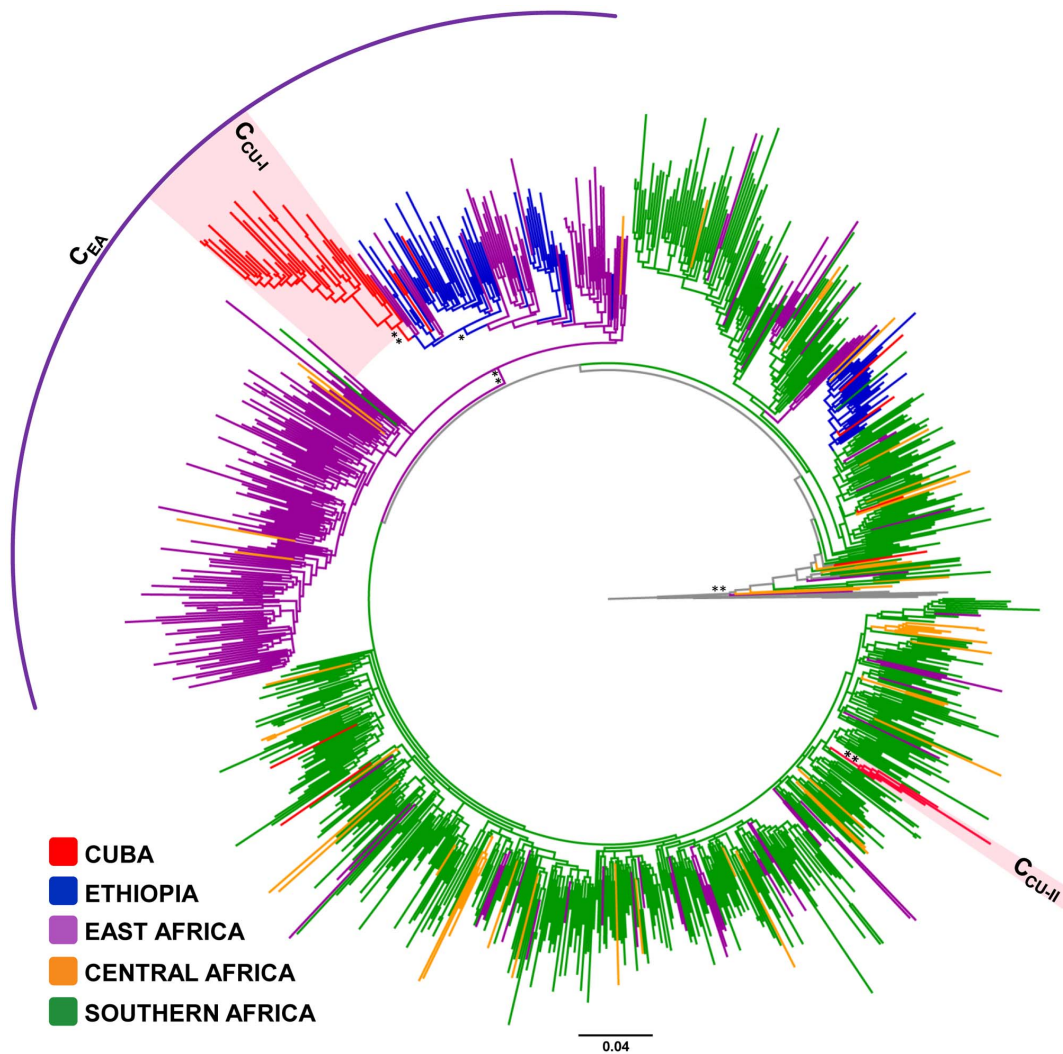


Figure 2. ML phylogenetic tree of HIV-1 subtype C *pol* (~1000 pb) sequences circulating in Cuba ($n=49$), and in central ($n=53$), eastern ($n=330$) and southern ($n=545$) African countries. Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). The circular bracket highlights the position of the subtype C east African clade (C_{EA}). Shaded boxes highlight the position of the two major HIV-1 subtype C Cuban clades (C_{CU-I} and C_{CU-II}). Key nodes with $aLRT$ support values >0.80 (*) and ≥ 0.90 (**) are indicated. The tree was rooted using HIV-1 subtype A1 and D reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. doi:10.1371/journal.pone.0072448.g002

CRF18_cpx ($n=15$) and CRFs20/23/24_BG ($n=7$) sequences from other countries (Tables S4 and S5).

Subtype assignment

The subtype assignment and recombinant structure of all sequences here included was confirmed by: REGA HIV subtyping tool v.2 [22]; Bootscanning with Simplot software v3.5.1 [23] and Maximum Likelihood (ML) phylogenetic analysis. In bootscan analyses, supporting branching with reference sequences from all HIV-1 group M subtypes were determined in Neighbor-Joining trees based on 100 re-samplings, within a 250 bp window moving in steps of 10 bases. ML phylogenetic trees were inferred under the best nucleotide substitution model selected using the jModeltest program [24] (Table S6). The ML tree was reconstructed with the PhyML program [25] using an online web server [26]. Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test ($aLRT$) [27]

based on the Shimodaira-Hasegawa-like procedure. The ML trees were visualized using the FigTree v1.4.0 program [28]. All HIV-1 sequences displaying incorrect clade assignment and/or frameshift mutations were excluded from the study, with the exception of four CRF23_BG sequences that were reclassified as CRF20_BG (GenBank accession numbers FJ481689 and FJ585687) and CRF24_BG (GenBank accession numbers JQ585465 and FJ481688).

Reconstruction of evolutionary 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), and the mode and rate (r , years $^{-1}$) of population growth of different Cuban HIV-1 clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.7.5 [29,30]. Analyses were performed using the best nucleotide substitution model (Table S6) and an uncorrelated Lognormal relaxed

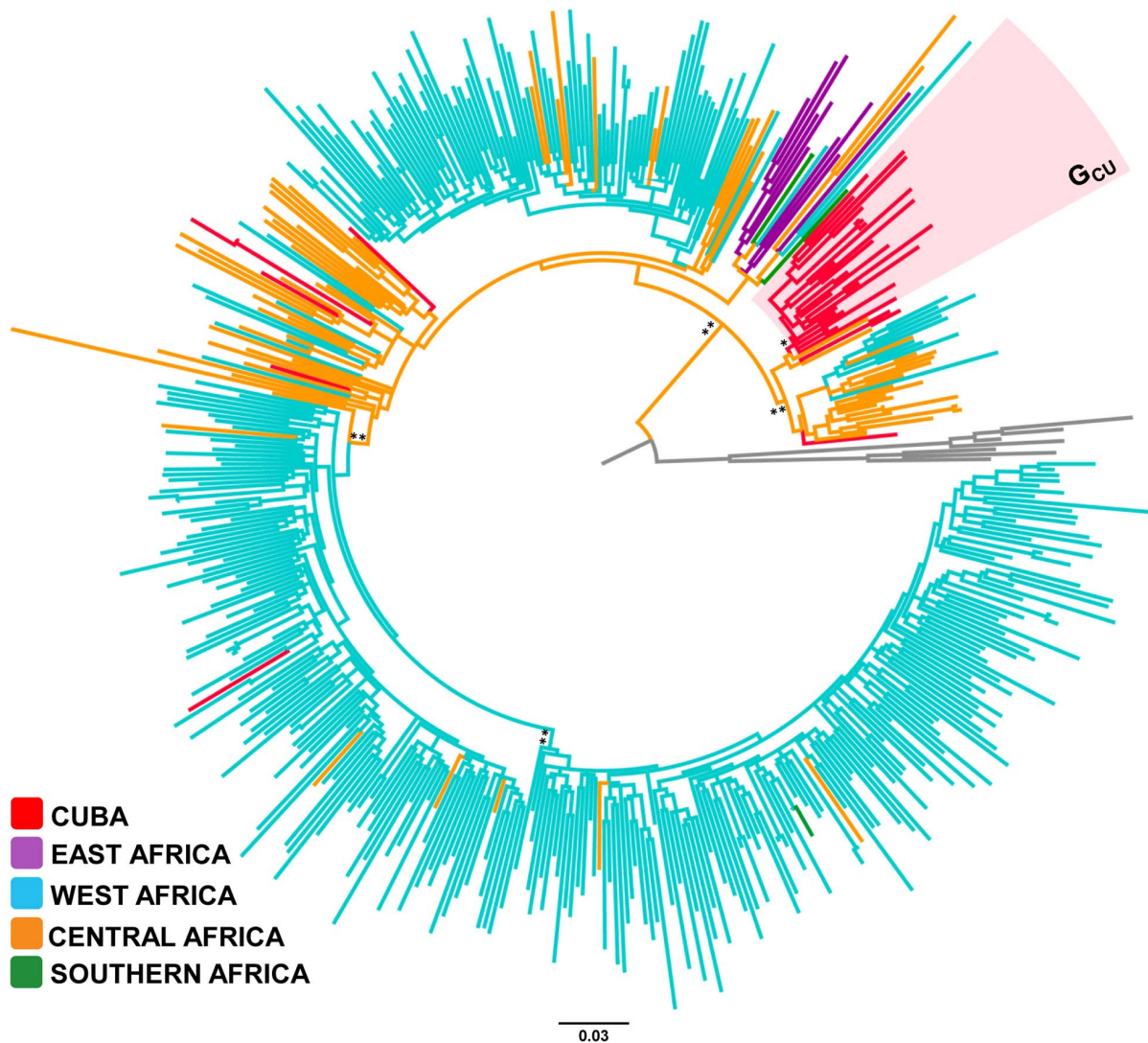


Figure 3. ML phylogenetic tree of HIV-1 subtype G *pol* (~1000 pb) sequences circulating in Cuba ($n=35$), and in central ($n=71$), western ($n=366$), eastern ($n=10$) and southern ($n=3$) African countries. Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). Shaded boxes highlight the position of the major HIV-1 subtype G Cuban clade (G_{CU}). Key nodes with $aLRT$ support values >0.80 (*) and ≥ 0.90 (**) are indicated. The tree was rooted using HIV-1 subtype A1 and B reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. doi:10.1371/journal.pone.0072448.g003

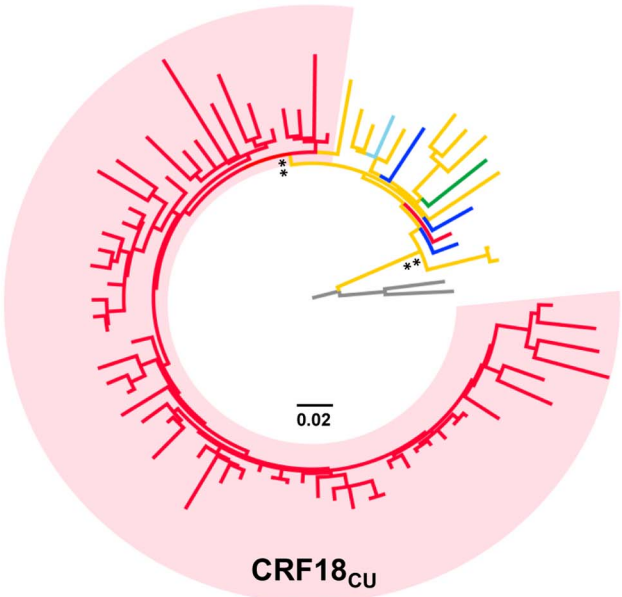
molecular clock model [31]. A Bayesian Skyline coalescent tree prior [32] was first used to estimate μ , the T_{mrea} , and the change in effective population size through time. Estimates of the population growth rate were subsequently obtained using a logistic growth coalescent tree prior that was the model pointed out by the Bayesian Skyline plot and that also provided the best fit to the demographic signal contained in most 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 [33]. MCMC chains were run for $10\text{--}50 \times 10^6$ 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, after excluding an initial 10% using the TRACER v1.5 program [34].

Results

Identification of major HIV-1 Cuban clades

The ML analysis of HIV-1 subtypes B, C and G sequences from Cuba and other countries from the Americas, Europe and Africa revealed that most Cuban strains branched in well-supported country-specific sub-clades. Of the 322 HIV-1 subtype B Cuban sequences analyzed, 180 (56%) formed a large country-specific monophyletic sub-clade (B_{CU-I} , $aLRT=0.81$), 44 (14%) branched in two clusters of medium size ($15 < n < 30$), 52 (16%) formed clusters of small size ($n \leq 10$), and the remaining 46 (14%) represented non-clustered sequences (Fig. 1). Of note, all subtype B Cuban sequences branched in a large $B_{PANDEMIC}$ monophyletic cluster ($aLRT=0.80$) together with most subtype B sequences from US (92%) and all sequences from France (100%); whereas most non-Cuban Caribbean sequences (60%) occupy the deepest branches within B clade (Fig. 1).

A)



B)

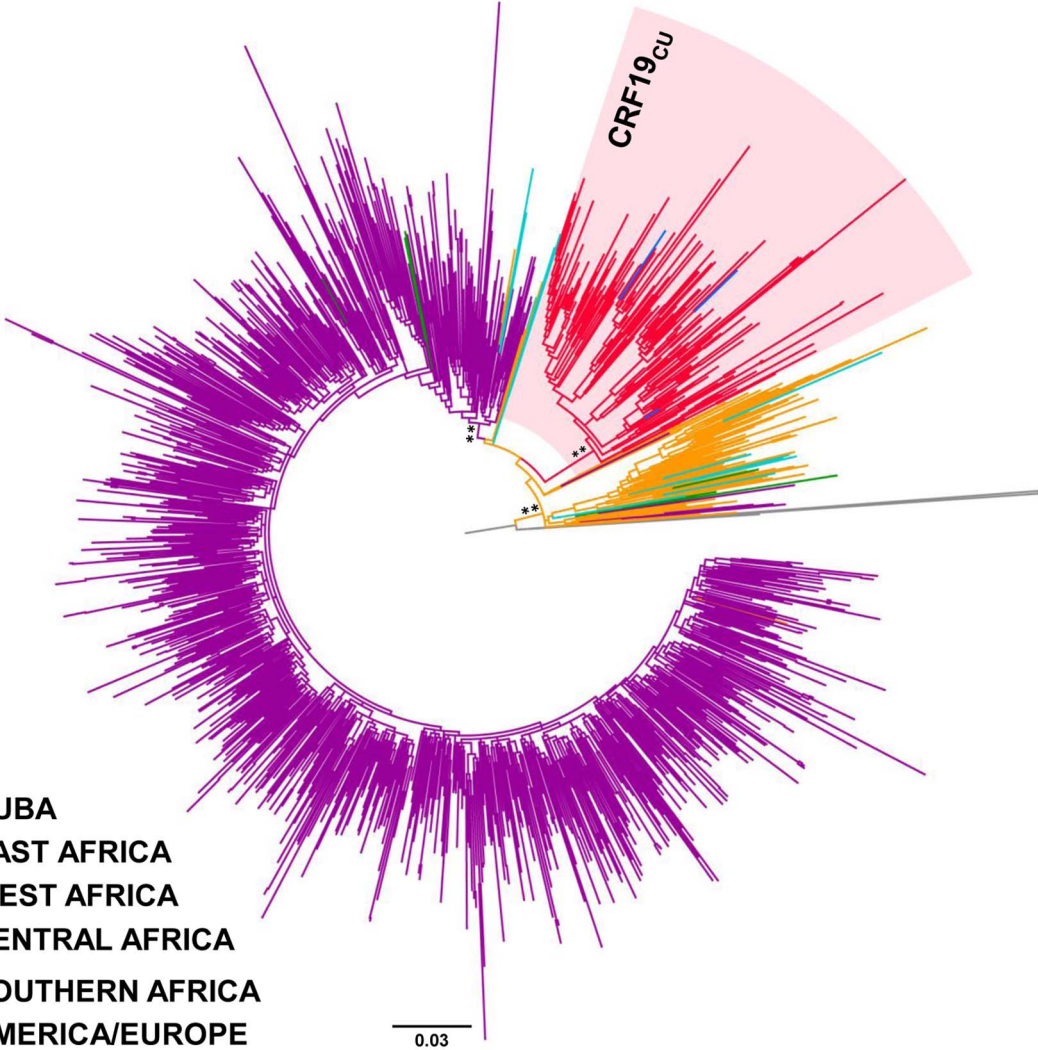


Figure 4. ML phylogenetic trees of HIV-1 CRFs_cpx pol (~1000 pb) sequences. A) HIV-1 CRF18_cpx from Cuba ($n=62$), were combined with those isolated in African ($n=12$), American ($n=1$) and European ($n=2$) countries. The tree was rooted using HIV-1 subtype G reference sequences (black branches). B) HIV-1 CRF19_cpx sequences from Cuba ($n=160$) and European countries ($n=3$) were combined with subtype D sequences of African origin ($n=1,112$). Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). Shaded boxes highlight the position of the major HIV-1 CRF18_cpx (CRF18_{CU}) and CRF19_cpx (CRF19_{CU}) Key nodes with $aLRT$ support values >0.80 (*) and ≥ 0.90 (**) are indicated. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. doi:10.1371/journal.pone.0072448.g004

Of the 49 HIV-1 subtype C Cuban sequences analyzed, 34 (69%) branched in a single monophyletic sub-cluster (C_{CU-I}, $aLRT=0.94$), six (12%) branched in a second well supported minor clade (C_{CU-II}, $aLRT=0.98$), and the remaining nine (18%) represented non-clustered sporadic lineages (Fig. 2). The major clade C_{CU-I} was nested within Ethiopian sequences that belongs to the previously called C_{EA} clade [35], a viral lineage characteristic of the east African region (Fig. 2). The minor clade C_{CU-II}, by contrast, was nested within subtype C sequences from southern

Africa (Fig. 2). Non-clustered Cuban sequences were scattered among strains from Ethiopia and southern African countries.

Of the 35 HIV-1 subtype G Cuba sequences analyzed, 26 (74%) branched in a single monophyletic sub-cluster (G_{CU}, $aLRT=0.87$) and the remaining nine (26%) represented sporadic lineages of one or two sequences. Although most subtype G African strains included in our analysis were from the western region ($n=366$, 84%), the clade G_{CU} and most sporadic subtype G Cuban lineages were nested among basal sequences from the central African region (Angola, DRC and Cameroon) (Fig. 3). There was only one

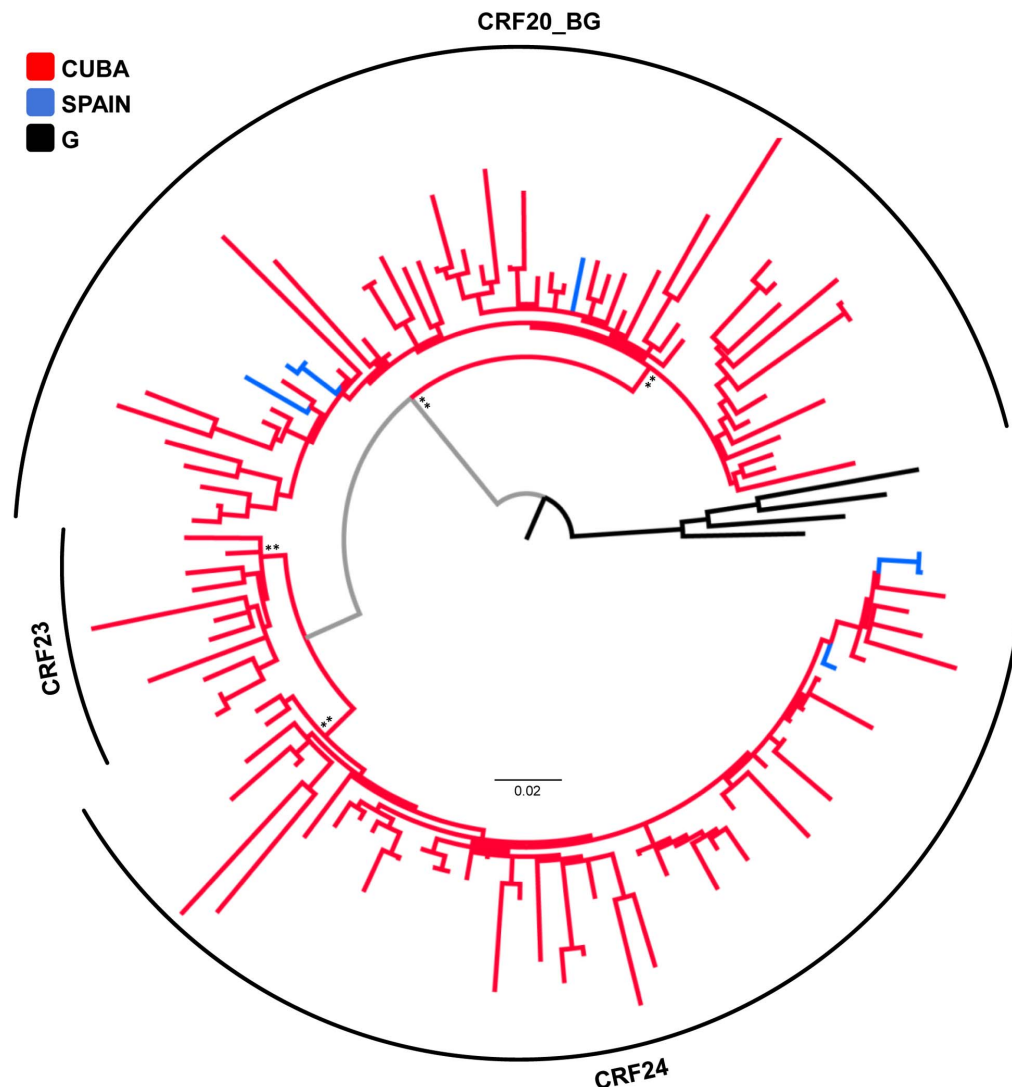


Figure 5. ML phylogenetic tree of HIV-1 CRFs20/23/24_BG pol (~1000 pb) sequences circulating in Cuba ($n=118$) and Spain ($n=7$). Branches are colored according to the origin of each sequence, as indicated at the legend (top left). The circular brackets highlight the distribution of the three CRFs_BG clades. The tree was rooted using HIV-1 subtype G reference sequences (black branches). Key nodes with $aLRT$ support values >0.80 (*) and ≥ 0.90 (**) are indicated. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. doi:10.1371/journal.pone.0072448.g005

Table 1. Evolutionary rate and time-scale of major HIV-1 Cuban clades.

HIV-1 clade	N	Sampling interval	μ (subst./site/year)	Coefficient of variation	T_{MRCA}
B _{CU-I}	176	2003–2011	3.0×10^{-3}	0.30	1992
			(2.4×10^{-3} – 3.6×10^{-3})	(0.21–0.39)	(1988–1994)
B _{CU-II}	27	1999–2011	2.4×10^{-3}	0.25	1991
			(1.6×10^{-3} – 3.2×10^{-3})	(0.01–0.45)	(1986–1994)
C _{CU}	34	2003–2011	2.8×10^{-3}	0.41	1994
			(2.0×10^{-3} – 3.8×10^{-3})	(0.19–0.65)	(1990–1998)
G _{CU}	26	1999–2011	2.0×10^{-3}	0.56	1988
			(1.0×10^{-3} – 3.3×10^{-3})	(0.36–0.81)	(1976–1995)
CRF18 _{CU}	61	1999–2011	2.6×10^{-3}	0.40	1992
			(1.9×10^{-3} – 3.5×10^{-3})	(0.25–0.59)	(1987–1996)
CRF19 _{CU}	158	1999–2011	3.4×10^{-3}	0.38	1987
			(2.9×10^{-3} – 4.0×10^{-3})	(0.30–0.47)	(1983–1991)
CRF20/23/24_BG	117	1999–2011	2.6×10^{-3}	0.35	1991
			(2.1×10^{-3} – 3.1×10^{-3})*	(0.25–0.45)*	(1986–1994)*
CRF20_BG	56	1999–2011	2.6×10^{-3}	0.35	1996
			(2.1×10^{-3} – 3.1×10^{-3})*	(0.25–0.45)*	(1994–1998)*
			2.4×10^{-3}	0.27	1996
CRF23_BG	11	2003–2011	2.6×10^{-3}	0.35	1998
			(2.1×10^{-3} – 3.1×10^{-3})*	(0.25–0.45)*	(1996–2000)*
CRF24_BG	50	2003–2011	2.6×10^{-3}	0.35	1997
			(2.1×10^{-3} – 3.1×10^{-3})*	(0.25–0.45)*	(1996–1999)*
			2.2×10^{-3}	0.36	1998
			(1.6×10^{-3} – 2.8×10^{-3})	(0.19–0.54)	(1996–2000)

*Estimates obtained from the combined CRF20/23/24_BG data set.
doi:10.1371/journal.pone.0072448.t001

Cuban sequence that branched within a major African subtype G sub-clade mainly composed by sequences from Nigeria.

Test the monophyletic origin of the HIV-1 CRFs_{cpx} Cuban sequences was very much complicated because the scarcity of CRF18_{cpx} ($n = 12$) and the absence of CRF19_{cpx} *pol* sequences of African origin available in public databases. Because CRF19_{cpx} is subtype D in the *pol* fragment analyzed, we decided to include all available subtype D *pol* sequences of African origin in our analysis. ML analysis revealed that all (except one) CRF18_{cpx} and all CRF19_{cpx} sequences from Cuba branched in highly supported ($aLRT \geq 0.90$) monophyletic sub-clusters (CRF18_{CU} and CRF19_{CU}) that were nested within CRF18_{cpx} and subtype D *pol* sequences of central African origin, respectively (Fig. 4). The few CRF18_{cpx} isolated in Europe ($n = 2$) and South America ($n = 1$) were intermixed among basal African strains; whereas all CRF19_{cpx} detected in Europe ($n = 3$) branched within the clade CRF19_{CU} (Fig. 4).

As expected, the CRFs20/23/24_BG Cuban sequences formed three well-supported ($aLRT \geq 0.90$) monophyletic lineages (Fig. 5). The few CRF20_BG ($n = 4$) and CRF24_BG ($n = 3$) sequences isolated outside Cuba (Spain and Greece) were intermixed among Cuban strains (Fig. 5); thus supporting a Cuban origin for all those European sequences.

Time scale of major HIV-1 Cuban clades

Bayesian MCMC analyses under a relaxed molecular clock model were used to estimate the substitution rate and T_{MRCA} of all

HIV-1 Cuban clades with a minimum size of 25 sequences. A few subtype B ($n = 4$) and CRF19_{cpx} ($n = 2$) sequences with anomalously long branches in the phylogenetic tree, were excluded. The final number of HIV-1 Cuban sequences included in the evolutionary analyses is shown in Table 1. The median estimated evolutionary rates for the *pol* region of the different HIV-1 clades were roughly similar, ranging from 2.0×10^{-3} subst./site/year (G_{CU} clade) to 3.4×10^{-3} subst./site/year (CRF19_{CU} clade), with a considerable overlap of the 95% HPD intervals (Table 1). The coefficient of rate variation was higher than zero for all HIV-1 datasets analyzed (Table 1), thus supporting the use of a relaxed molecular clock model to reconstruct the time-scale of major HIV-1 Cuban lineages.

The median T_{MRCA} of those HIV-1 clades imported into Cuba range between 1987 (CRF19_{CU}) and 1994 (C_{CU-I}); whereas the median T_{MRCA} of those CRF20_BG variants locally generated varied between 1996 and 1998 (Table 1). The T_{MRCA} of CRF20_BG and CRF24_BG clades estimated from the single CRF datasets were almost identical to those estimated from the combined CRFs20/23/24_BG data set (Table 1), indicating that all Cuban CRFs_{BG} evolved at quite similar rates. A previous study [20], proposed that Cuban CRF_{BG} viruses derive from a common recombinant ancestor generated by recombination between clade G_{CU} and the second most prevalent subtype B clade (B_{CU-II}) (Fig. 1). The analysis of the combined CRFs20/23/24_BG data set allows us to estimate the median T_{MRCA} of that putative BG recombinant ancestor at 1991, roughly coinciding

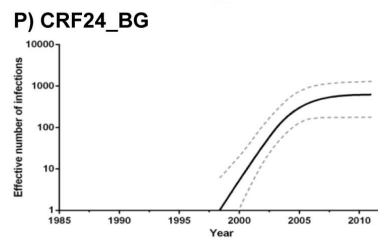
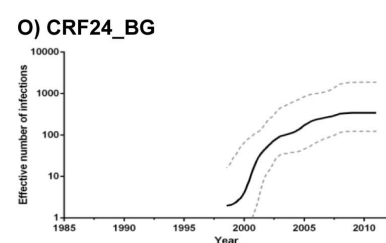
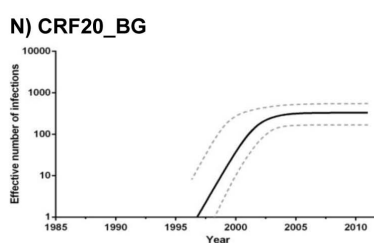
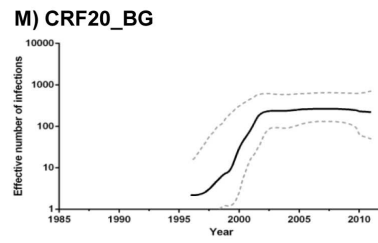
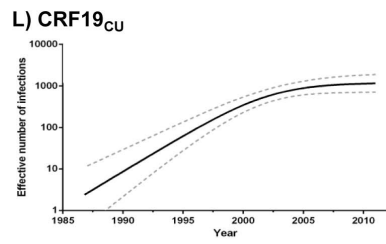
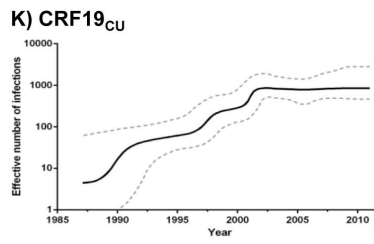
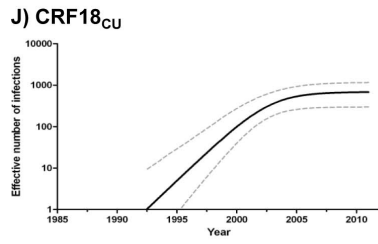
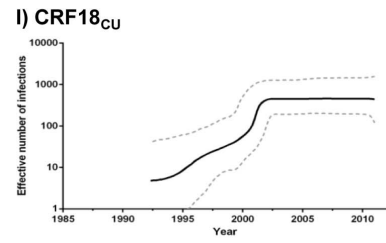
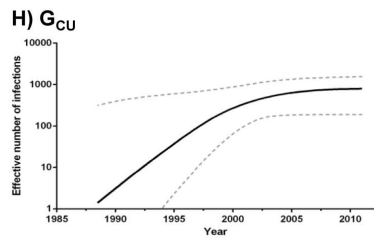
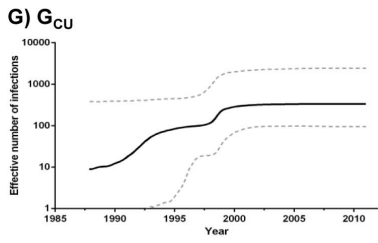
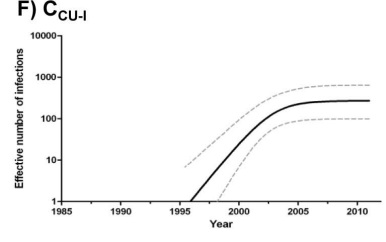
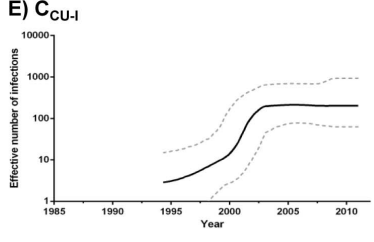
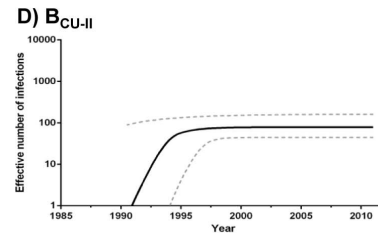
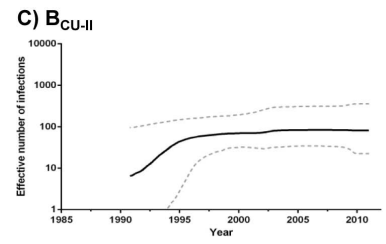
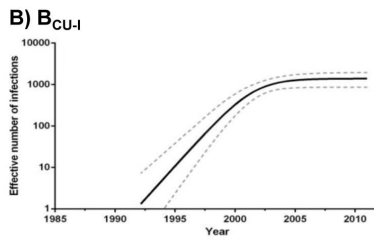
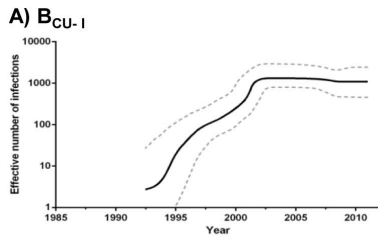


Figure 6. Demographic history of the major HIV-1 Cuban clades. Effective number of infections through time estimated using both Bayesian skyline (A, C, E, G, I, K, M and O) and logistic growth (B, D, F, H, J, L, N and P) coalescent models are shown for each of the eight HIV-1 Cuban clades analyzed. Median estimates of the effective number of infections (solid line) and 95% HPD intervals of the estimates (dashed lines) are shown in each graphic. The vertical axes represent the estimated effective number of infections on a logarithmic scale. Time scale is in calendar years. doi:10.1371/journal.pone.0072448.g006

with the estimated T_{MRCA} of the parental clades G_{CU} and B_{CU-II} (Table 1).

Demographic history of major HIV-1 Cuban clades

The Bayesian skyline plot (BSP) analyses suggest that all HIV-1 Cuban clades experienced an initial phase of fast exponential growth followed by a more recent decline in growth rate (Fig. 6). The growth rate of most HIV-1 Cuban clades seems to start to decrease around the early 2000s; except for clades B_{CU-II} and CRF24_BG that seem to stabilize at earlier (before 2000) and later (after 2005) time points, respectively. The BSP analyses also suggests that the coalescent model of logistic population growth fits the demographic information contained in all HIV-1 Cuban data sets better than the exponential one.

To test this, log ML for the logistic and exponential growth models were calculated using both PS and SS methods. The model of logistic population growth was strongly supported over the exponential one for most HIV-1 Cuban clades ($\log BF > 3$), with exception of G_{CU} and CRF24_BG for which only a weak support was obtained ($\log BF = 0.9-1.0$) (Table 2). Such a low BF support to the logistic growth model could be explained by the low number of sequences in clade G_{CU} ($n = 26$) and the very recent stabilization of clade CRF24_BG (after 2005). Moreover, the overall time-scale and demographic pattern obtained from both BSP and logistic growth coalescent tree priors were very similar for all HIV-1 Cuban clades (Fig. 6). According to the logistic model, the median initial growth rates of HIV-1 Cuban clades range between 0.40 year⁻¹ (CRF19_{CU}) to 1.57 year⁻¹ (B_{CU-II}) with some overlap of the 95% HPD intervals for most lineages, except between

CRF19_{CU} and clades B_{CU-I} , B_{CU-II} , C_{CU-I} , CRF20_BG and CRF24_BG (Fig. 7).

Discussion

The Cuban HIV epidemic is unique in the Americas because of the exceptionally low HIV prevalence, estimated at 0.20% in adults in 2011 [36], and the unusually high HIV-1 genetic diversity with circulation of subtype B and several non-B subtypes [2,3,4,5,6,7]. Our study indicates that most HIV-1 infections in Cuba derived from the dissemination of a few founder viruses that were either introduced from the Americas/Europe (subtype B) and Africa (subtype C, subtype G, CRF18_cpx and CRF19_cpx) or were locally generated (CRFs20/23/24_BG).

The most accepted model of worldwide HIV-1 subtype B dissemination suggests that the virus moved from Haiti to other Caribbean islands and to the United States (US), and then from the US to the rest of the world establishing a “B_{PANDEMIC}” clade [37]. The phylogenetic analysis here performed revealed multiple ($n \geq 66$) introductions of HIV-1 B_{PANDEMIC} strains in Cuba, although the bulk of the subtype B epidemic in this country resulted from the dissemination of only a few clades. The two most prevalent clades B_{CU-I} and B_{CU-II} comprises about 55% and 8% of all subtype B sequences from Cuba here included, respectively. We estimate that these clades most probably emerged in Cuba in the early 1990s, much later than the estimated origin of subtype B epidemics in Haiti and the US (1960–1970) [37,38,39]. The estimated T_{MRCA} of clades B_{CU-I} and B_{CU-II} coincides with a crisis in the Cuban economy caused by the collapse of the Soviet Union in 1991 that precipitated important investments in the tourist industry and a sharp increase in the number of tourist mostly from

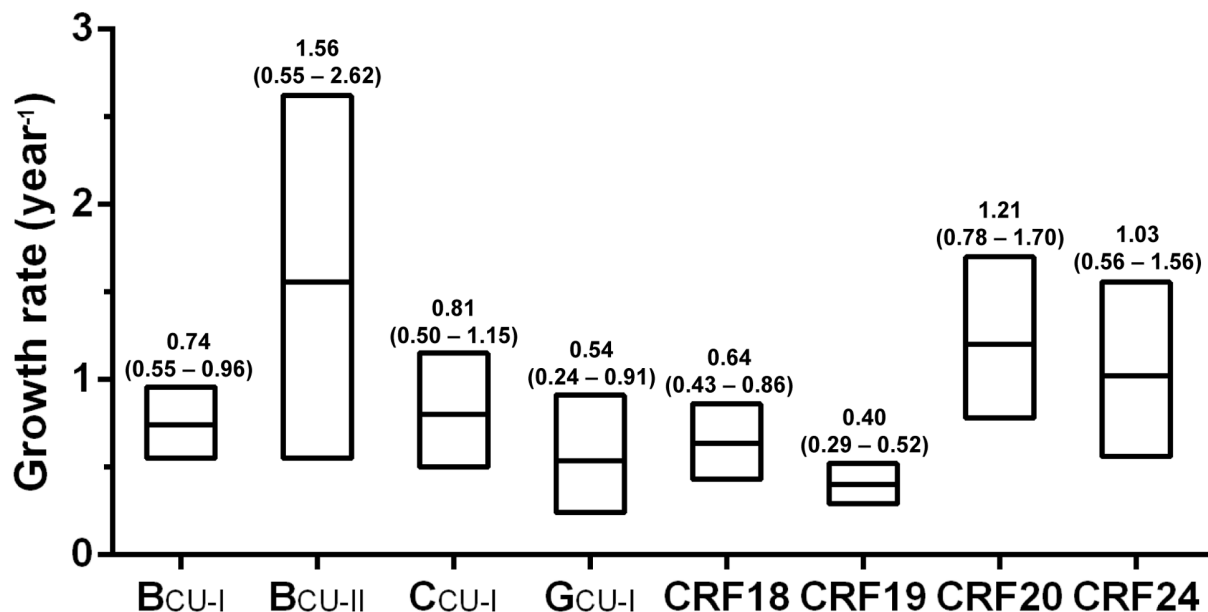


Figure 7. Coalescent estimates of epidemic growth rate of the major HIV-1 Cuban clades. The box plots and the numbers above represent the median growth rates (years⁻¹) and the 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each of the eight HIV-1 Cuban clades analyzed. doi:10.1371/journal.pone.0072448.g007

Table 2. Best fit demographic model for major HIV-1 Cuban clades.

Dataset	PS			SS		
	Log ML LG	Log ML EG	Log BF (LG vs EG)	Log ML LG	Log ML EG	Log BF (LG vs EG)
B _{CUBA I}	-12891.71	-12935.39	43.68	-12894.56	-12939.47	43.91
B _{CUBA II}	-3563.92	-3568.51	4.59	-3563.99	-3568.71	4.72
C _{CUBA}	-4114.63	-4118.50	3.87	-4114.74	-4118.72	3.98
G _{CUBA}	-4196.64	-4197.66	1.02	-4196.87	-4197.73	0.86
CRF18	-6109.29	-6118.91	9.62	-6109.54	6119.01	9.47
CRF19	-14552.57	-14564.21	11.64	-14553.91	-14565.09	11.18
CRF20	-5083.74	-5100.80	17.06	-5083.91	-5101.05	17.14
CRF24	-4159.40	-4160.43	1.03	-4159.81	-4160.70	0.89

Log marginal likelihood (ML) estimates for logistic growth (LG) and exponential growth (EG) 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 = LG) and null (H0 = EG) models. Log BF_s>1 indicates that model H1 is more strongly supported by the data than model H0.

doi:10.1371/journal.pone.0072448.t002

North America and Europe [40], regions with a widespread circulation of the subtype B_{PANDEMIC} clade. This may explain the massive influx of subtype B_{PANDEMIC} strains and the apparent absence of “non-pandemic” subtype B Caribbean lineages in Cuba.

Similarly to subtype B, there were multiple introductions of subtype C ($n \geq 10$), subtype G ($n \geq 8$) and CRF18_{cpx} ($n \geq 2$) viruses in Cuba, but only a few of them were able to get established and disseminate. The clades C_{CU-I}, G_{CU} and CRF18_{CU} comprise 69%, 74% and 98% of all subtype C, subtype G and CRF18_{cpx} sequences from Cuba included in this study, respectively. The monophyletic clustering of CRF19_{cpx}-like *pol* Cuban sequences within subtype D radiation, the paucity of this genetic variant in Africa, and the recent T_{MRC}A of Cuban sequences strongly suggests that the CRF19_{CU} clade also derives from a single founder event. HIV-1 clades G_{CU}, CRF18_{CU} and CRF19_{CU} probably originate in central Africa, whereas clade C_{CU-I} probably derives from east Africa. Our study suggests that clades CRF19_{CU} and G_{CU} began to circulate in Cuba around the late 1980s, followed shortly thereafter by clades CRF18_{CU} and C_{CU-I}. Thus, although Cuban personnel were stationed in several African countries since the 1970s, HIV-1 African strains were successfully disseminated within Cuba only from the late 1980s onwards.

Our data suggest that HIV-1 CRFs_{BG} (20_{BG}, 23_{BG} and 24_{BG}) started to spread in Cuba in the second half of the 1990s. Such a recent expansion of BG recombinants in Cuba is fully consistent with epidemiological data showing that in samples collected in 2003, none of the individuals harboring BG recombinants were diagnosed with HIV-1 infection earlier than 1996, and all but three were diagnosed since 2000 [20]. Similarly, the proportion of BG infections among MSM in Havana City increased from 0% in those diagnosed in 1998 to 31% in those diagnosed in 2003 [3]. It was proposed that all Cuban CRFs_{BG} evolved from a common BG recombinant ancestor locally generated by recombination between parental clades B_{CU-II} and G_{CU} [20]. According to our estimations, that common BG recombinant ancestor was generated in the early 1990s, thus around or immediately after the estimated onset date of parental clades B_{CU-II} and G_{CU} and some years earlier than the emergence of the CRFs_{BG}.

The reconstruction of the demographic history indicates that most HIV-1 Cuban clades followed a very similar growth pattern characterized by rapid dissemination until the early 2000s after which the epidemic growth rate of those epidemics started to slow-down. The expansion of the B_{CU-II} clade, by contrast, seems to decrease during 1990s; whereas the growth rate of the CRF24_{BG} clade probably only stabilized in the second half of the 2000s. The initial expansion of the major HIV-1 Cuban clades coincides with a sustained increase in the number of infected HIV-positive individuals in Cuba from 1991 to 2000 [41]. UNAIDS estimations indicate that the total number of people living with HIV in Cuba continued to grow in the last decade, rising from 3,100 (2,600–4,300) in 2000 to 14,000 (12,000–16,000) in 2011 [36]. Our demographic analysis, however, suggests a trend toward stability in the effective number of infections of all major HIV-1 Cuban clades over time consistent with recent epidemiological data that shows a decrease of HIV incidence in Cuba, mainly among men, in the biennium 2010–2011 [42].

Our coalescent-based analyses suggest that CRF20_{BG}, CRF24_{BG} and B_{CU-II} have displayed a more explosive initial growth (1.0 year⁻¹–1.6 year⁻¹) than clades G_{CU}, CRF18_{CU} and CRF19_{CU} (~0.4–0.6 year⁻¹); whereas clades B_{CU-I} and C_{CU-I} displayed intermediate initial growth rates (~0.8 year⁻¹). Notably, all those HIV-1 Cuban clades with the fastest initial expansion rates (B_{CU-I}, B_{CU-II}, B_{CU-I} and CRFs_{BG}) were much more prevalent among MSM than among heterosexual (HET) persons [3]. Thus, some HIV-1 Cuban clades may have spread faster than others because they encountered, by chance, local transmission chains with higher rates of partner exchange. Dissemination within a transmission network of small size and high rate of partner exchange may also explain the fast, but self limited, dissemination phase of clade B_{CU-II}. Additional influence of virological factors cannot be excluded. These results must also be interpreted with caution as most growth rate estimates here obtained displayed quite large overlapping 95% HPD intervals.

It has been proposed that Cuba's low rate of HIV infection is due to several factors that served to prevent sexual transmission of the virus, including: wide-scale HIV screening and subsequent contact tracing of HIV-positive individuals, mandatory quarantine of the first HIV-infected individuals at sanatoria, free access to a well structured public health system, comprehensive HIV education campaigns, coordinated work of Cuban government agencies

and community, and restricted tourism between Cuba and western countries up to the early 1990s [43,44,45,46]. The estimated initial growth rates of the major HIV-1 Cuban clades ($\sim 0.4\text{--}1.6\text{ year}^{-1}$), however, were comparable to those obtained for different HIV-1 epidemics in the Americas ($\sim 0.5\text{--}1.3\text{ year}^{-1}$) [38,47,48,49,50,51,52], Europe ($\sim 0.4\text{--}1.5\text{ year}^{-1}$) [52,53,54,55], Africa ($\sim 0.2\text{--}0.8\text{ year}^{-1}$) [47,52,56,57,58] and Asia ($\sim 0.8\text{ year}^{-1}$) [59]. This suggests that several factors may have contributed to delay the introduction and/or dissemination of HIV-1 in Cuba for many years; but once some HIV-1 strains got established in vulnerable HET and MSM transmission groups they spread quickly.

In summary, this study indicates that only a few subtype B and non-B subtype founder viral strains were successfully disseminated in Cuba. Some of those HIV-1 viral strains were probably introduced from North America/Europe, central Africa and east Africa between the middle 1980s and the middle 1990s; whereas other were locally generated around the late 1990s. Changes in the social and economic landscapes of Cuba occurring at the beginning of the 1990s may have fueled the introduction and/or initial dissemination of major HIV-1 Cuban clades. Although the main HIV-1 Cuban lineages began to circulate at a rather late time of the AIDS pandemic, further dissemination within vulnerable groups was rapid. These results reinforce the importance of maintaining, reviewing and updating permanently the public health measures aimed at controlling the spread of those HIV-1 variants already established in the Cuban population.

References

- Thomson MM, Najera R (2005) Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev* 7: 210–224.
- Cuevas MT, Ruibal I, Villahermosa ML, Diaz H, Delgado E, et al. (2002) High HIV-1 genetic diversity in Cuba. *Aids* 16: 1643–1653.
- Perez L, Thomson MM, Bleda MJ, Aragonés C, Gonzalez Z, et al. (2006) HIV Type 1 molecular epidemiology in Cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. *AIDS Res Hum Retroviruses* 22: 724–733.
- Kouri V, Aleman Y, Perez L, Perez J, Fonseca C, et al. (2012) High frequency of antiviral drug resistance and non-B subtypes in HIV-1 patients failing antiviral therapy in Cuba. *J Clin Virol* 55: 348–355.
- Machado LY, Blanco M, Dubed M, Diaz HM, Ruiz NM, et al. (2012) HIV type 1 genetic diversity in newly diagnosed Cuban patients. *AIDS Res Hum Retroviruses* 28: 956–960.
- Machado LY, Dubed M, Diaz H, Ruiz N, Romay D, et al. (2013) Transmitted HIV type 1 drug resistance in newly diagnosed Cuban patients. *AIDS Res Hum Retroviruses* 29: 411–414.
- Perez L, Kouri V, Aleman Y, Abrahantes Y, Correa C, et al. (2013) Antiretroviral drug resistance in HIV-1 therapy-naive patients in Cuba. *Infect Genet Evol* 16C: 144–150.
- Thomson MM, Casado G, Posada D, Sierra M, Najera R (2005) Identification of a novel HIV-1 complex circulating recombinant form (CRF18_cpx) of Central African origin in Cuba. *Aids* 19: 1155–1163.
- Casado G, Thomson MM, Sierra M, Najera R (2005) Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. *J Acquir Immune Defic Syndr* 40: 532–537.
- Bartolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, 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.
- 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.
- Kalish ML, Robbins KE, Pieniazek D, Schaefer A, Nzilambi N, et al. (2004) Recombinant viruses and early global HIV-1 epidemic. *Emerg Infect Dis* 10: 1227–1234.
- Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, et al. (2000) Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 74: 10498–10507.
- Niama FR, Toure-Kane C, Vidal N, Obengui P, Bikandou B, et al. (2006) HIV-1 subtypes and recombinants in the Republic of Congo. *Infect Genet Evol*.
- Pircher M, Diafouka M, Papuchon J, Recordon-Pinson P, Mahambou DN, et al. (2012) Molecular characterization of HIV type 1 in Brazzaville, Republic of

Supporting Information

Table S1 HIV-1 subtype B dataset.
(PDF)

Table S2 HIV-1 subtype C dataset.
(PDF)

Table S3 HIV-1 subtype G dataset.
(PDF)

Table S4 HIV-1 CRF18_cpx and CRF19_cpx/subtype D datasets.
(PDF)

Table S5 HIV-1 CRF20/23/24_cpx datasets.
(PDF)

Table S6 Nucleotide substitution models selected using jModeltest program.
(PDF)

Acknowledgments

We wish to thank Dr Vera Bongertz for critical reviewing of the manuscript.

Author Contributions

Conceived and designed the experiments: GB ED. Performed the experiments: ED GB. Analyzed the data: ED GB. Wrote the paper: GB ED.

- Congo, and first data on resistance to antiretroviral drugs. *AIDS Res Hum Retroviruses* 28: 1798–1802.
- Marechal V, Jauvin V, Selekon B, Leal J, Pelembi P, et al. (2006) Increasing HIV type 1 polymorphic diversity but no resistance to antiretroviral drugs in untreated patients from Central African Republic: a 2005 study. *AIDS Res Hum Retroviruses* 22: 1036–1044.
- Carr JK, Torimiro JN, Wolfe ND, Eitel MN, Kim B, et al. (2001) The AG recombinant IBNG and novel strains of group M HIV-1 are common in Cameroon. *Virology* 286: 168–181.
- Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnotou R, Mbuagbaw J, et al. (2003) Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage. *Virology* 310: 254–266.
- Powell R, Barengolts D, Mayr L, Nyambi P (2010) The Evolution of HIV-1 Diversity in Rural Cameroon and its Implications in Vaccine Design and Trials. *Viruses* 2: 639–654.
- Sierra M, Thomson MM, Posada D, Perez L, Aragonés C, et al. (2007) Identification of 3 phylogenetically related HIV-1 BG intersubtype circulating recombinant forms in Cuba. *J Acquir Immune Defic Syndr* 45: 151–160.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
- de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21: 3797–3800.
- Ray SC (1999) Simplot v2.5.0. Available: <http://sray.med.som.jhmi.edu/SCRsoftware/simplot/>. Accessed 2013 July 16.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253–1256.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
- 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.
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55: 539–552.
- Rambaut A (2009) FigTree v1.4.0: Tree Figure Drawing Tool. Available: <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 2013 July 16.
- 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.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.

31. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88.
32. 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.
33. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29: 2157–2167.
34. Rambaut A, Drummond A (2007) Tracer v1.4. Available: <http://tree.bio.ed.ac.uk/software/tracer/>. Accessed 2013 July 16.
35. Delatorre EO, Bello G (2012) Phylogenetics of HIV-1 Subtype C Epidemic in East Africa. *PLoS ONE* 7: e41904.
36. UNAIDS (2013) AIDS INFO DATABASE. Available: <http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/>. Accessed 2013 July 16.
37. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, et al. (2007) The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* 104: 18566–18570.
38. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, et al. (2003) U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol* 77: 6359–6366.
39. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, et al. (2011) Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS ONE* 6: e27489.
40. Figueras MA (2001) International Tourism and the Formation of Productive Clusters in the Cuban Economy. Latin American Studies Association (LASA) XXII Congress, Washington, DC Available: http://www3.dogus.edu.tr/memek/DersNotlari/TOUR%20229/TOUR%20229-2004%20fall/cuba_contribution.pdf. Accessed 2013 July 16.
41. Hsieh YH, Chen CW, Lee SM, de Arazoza H (2001) On the recent sharp increase in HIV detections in Cuba. *AIDS* 15: 426–428.
42. UNAIDS (2012) Country progress reports. Available: [http://www.unaids.org/en/dataanalysis/knownyourresponse/countryprogressreports/2012countries/cc_CU_Narrative_Report\[1\].pdf](http://www.unaids.org/en/dataanalysis/knownyourresponse/countryprogressreports/2012countries/cc_CU_Narrative_Report[1].pdf). Accessed 2013 July 16.
43. Perez-Stable EJ (1991) Cuba's response to the HIV epidemic. *Am J Public Health* 81: 563–567.
44. Hansen H, Groce NE (2001) From quarantine to condoms: shifting policies and problems of HIV control in Cuba. *Med Anthropol* 19: 259–292.
45. de Arazoza H, Joanes J, Lounes R, Legeai C, Clemencón S, et al. (2007) The HIV/AIDS epidemic in Cuba: description and tentative explanation of its low HIV prevalence. *BMC Infect Dis* 7: 130.
46. Anderson T (2009) HIV/AIDS in Cuba: lessons and challenges. *Rev Panam Salud Publica* 26: 78–86.
47. Salemi M, de Oliveira T, Soares MA, Pybus O, Dumans AT, et al. (2005) Different epidemic potentials of the HIV-1B and C subtypes. *J Mol Evol* 60: 598–605.
48. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimaraes ML, Chequer-Fernandez SL, et al. (2007) Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol* 7: 263–270.
49. Bello G, Guimaraes ML, Passaes CP, Matos Almeida SE, Veloso VG, et al. (2009) Short communication: Evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31_BC epidemics in southern Brazil. *AIDS Res Hum Retroviruses* 25: 1065–1069.
50. Bello G, Aulicino PC, Ruchansky D, Guimaraes ML, Lopez-Galindez C, et al. (2010) Phylogenetics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. *Retrovirology* 7: 22.
51. Ristic N, Zukurov J, Alkmim W, Diaz RS, Janini LM, et al. (2011) Analysis of the origin and evolutionary history of HIV-1 CRF28_BF and CRF29_BF reveals a decreasing prevalence in the AIDS epidemic of Brazil. *PLoS ONE* 6: e17485.
52. Bello G, Afonso JM, Morgado MG (2012) Phylogenetics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect Genet Evol* 12: 1079–1086.
53. Hue S, Pillay D, Clewley JP, Pybus OG (2005) Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci U S A* 102: 4425–4429.
54. Salemi M, de Oliveira T, Ciccozzi M, Rezza G, Goodenow MM (2008) High-Resolution Molecular Epidemiology and Evolutionary History of HIV-1 Subtypes in Albania. *PLoS ONE* 3: e1390.
55. Zehender G, Ebranati E, Lai A, Santoro MM, Alteri C, et al. (2010) Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J Acquir Immune Defic Syndr* 55: 156–160.
56. Walker PR, Pybus OG, Rambaut A, Holmes EC (2005) Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. *Infect Genet Evol* 5: 199–208.
57. Tully DC, Wood C (2010) Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. *Aids* 24: 1577–1582.
58. Delatorre E, Bello G (2013) Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa. *Aids*.
59. Chen JH, Wong KH, Chan KC, To SW, Chen Z, et al. (2011) Phylogenetics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS ONE* 6: e25286.