Title: Dissemination of non-pandemic Caribbean HIV-1 subtype B clades in Latin America

Short title: HIV-1 subtype B epidemic in Latin America

Authors: Marina CABELLO,[a] Dennis Maletich JUNQUEIRA,[b,c,d] and Gonzalo BELLO[a#]

Affiliation:

[a] Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil.

[b] Centro de Desenvolvimento Científico e Tecnológico, Fundação Estadual de Produção e Pesquisa em Saúde, Porto Alegre, RS, Brazil.

[c] Uniritter Laureate International Universities, Departamento de Ciências da Saúde, Porto Alegre, RS, Brazil.

[d] Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Universidade Federal do Rio Grande do Sul (UFGRS), Porto Alegre, RS, Brazil.

Total words in Abstract: 247

Total words in Manuscript: 3269

[#]**Corresponding author:**

Gonzalo Bello. Lab. de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ. Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil. Phone: +55 21 3865 8227. Fax: +55 21 3865 8173. E-mail: gbellobr@gmail.com / gbello@ioc.fiocruz.br

Conflicts of Interest and Source of Funding: The authors have no conflict of interest. This work was supported by Public Health Service grants E-26/110.439/2014 from the FAPERJ and 472896/2012-1 from the CNPq. MC was funded by a fellowship from Instituto Oswaldo Cruz -FIOCRUZ.

ABSTRACT

Objective: To estimate the prevalence of the HIV-1 subtype B pandemic ($B_{PANDEMIC}$) and Caribbean (B_{CAR}) clades in Latin America and to reconstruct the spatiotemporal dynamics of dissemination of the B_{CAR} clades in the region.

Design: A total of 7,654 HIV-1 subtype B *pol* sequences collected from 18 different Latin American countries between 1989 and 2011 were analyzed together with subtype B reference sequences representative of the $B_{PANDEMIC}$ (US/France = 300) and the B_{CAR} (Caribbean = 279, Panama = 37) clades.

Methods: Phylogeographic and evolutionary parameters were estimated from sequence data using Maximum Likelihood and Bayesian coalescent-based methods.

Results: Non-pandemic B_{CAR} strains were probably disseminated from the Caribbean islands of Hispaniola and Trinidad and Tobago into Latin America since the early 1970s. The B_{CAR} strains reached nearly all countries from Latin America here analyzed and in some of them were spread locally, although their overall prevalence in the region is low. The $B_{PANDEMIC}$ clade comprises >90% of subtype B infections in most countries analyzed, with exception of Suriname, French Guyana and probably Guyana, where both $B_{PANDEMIC}$ and B_{CAR} clades seem to circulate at a similar prevalence.

Conclusions: This study demonstrates that non-pandemic subtype B lineages of Caribbean origin have been disseminated into Latin America shortly after the estimated introduction of subtype B in the continent. Despite their early dissemination, the B_{CAR} strains account for a minor fraction of current HIV-1 subtype B infections in the region that are mainly driven by spreading of the globally disseminated $B_{PANDEMIC}$ clade.

Keywords: HIV-1; subtype B; non-pandemic; phylogeography; Latin America.

INTRODUCTION

An estimate of 1.5 million people were living with the Human Immunodeficiency Virus Type 1 (HIV-1) in Latin America at 2012, most of them concentrated in Brazil (40%), Mexico (11%), Colombia (10%), Venezuela (7%) and Argentina (6.5%) [1]. The HIV prevalence in the adult population (15-49 years) ranges from 0.2% in Mexico to >1.0% in Belize, Guyana and Suriname [1]. Most of the HIV epidemics in this region are concentrated in and around networks of men who have sex with men (MSM), although heterosexual HIV transmission is increasing in the older epidemics in South America and injecting drug use is another significant route of HIV transmission especially in the southern cone of South America and in Mexico [2].

The HIV-1 group M subtype B is the most prevalent clade in Latin America, accounting for about 70% of infections in the region [3]. The spread of HIV-1 subtype B in the Americas probably occurred via a single introduction event from Central Africa into Haiti around the middle 1960s and later dissemination of the virus from Haiti to other Caribbean islands and to the United States (US) [4]. The virus that entered the US was further disseminated from this country to other countries around the world, establishing a "$B_{PANDEMIC}$" clade, whereas other subtype B lineages seem to have remained mostly restricted to the Caribbean ("B_{CAR}" clades) [4]. A recent study conducted by our group analyzed 1,042 HIV-1 subtype B *pol* gene sequences from 14 different Caribbean countries and revealed that non-pandemic B_{CAR} lineages have been widely disseminated through the Caribbean region since the late 1960s, accounting for an important fraction of current HIV-1 infections in several countries including Haiti and the Dominican Republic (~75%), Jamaica (~50%) Trinidad and Tobago (~95%) and other Lesser Antilles (~40-75%) [5].

Two previous studies suggest that non-pandemic B_{CAR} lineages may have been also directly disseminated from the Caribbean islands into South [6] and Central [7] American countries. The study of Junqueira *et al* (2011) identified a few HIV-1 subtype B *pol* sequences from Brazil, Colombia, Guyana, Suriname and Venezuela that were phylogenetically intermixed among basal non-pandemic Caribbean sequences, suggesting a direct epidemiological link between the Caribbean and South American epidemics. Another recent study showed that a minor fraction (5.5%) of Panamanian subtype B *pol* sequences were also intermixed among non-pandemic B_{CAR} strains and further suggests that some of those B_{CAR} clades were mainly disseminated in Panama by heterosexual transmission [7]. Overall, these results suggest that the B_{CAR} clades have not remained confined to the Caribbean region; but have been also disseminated to continental regions of the Americas. The relative prevalence of the $B_{PANDEMIC}$ and B_{CAR} clades across different Latin American countries, however, remains largely unknown.

The objective of this study was to estimate the current prevalence of the $B_{PANDEMIC}$ and B_{CAR} clades in Latin America and to reconstruct the spatiotemporal dynamics of dissemination of the HIV-1 B_{CAR} clades in the region. For this, we used a comprehensive dataset of HIV-1 subtype B *pol* sequences (n = 7,654) isolated from 18 different Latin American countries between 1989 and 2011. These Latin American sequences were combined with subtype B reference sequences representative of the $B_{PANDEMIC}$ (US/France = 300) and the B_{CAR} (Caribbean/Panama = 316) clades and then subjected to Maximum Likelihood and Bayesian phylogeographic analyses.

METHODS

HIV-1 subtype B *pol* sequence dataset. We downloaded all HIV-1 subtype B *pol* sequences from Latin America that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nucleotides 2253–3260 relative to HXB2 clone) and were available at the Los Alamos HIV Database (http://www.hiv.lanl.gov) by December 2013. Additional HIV-1 subtype B *pol* sequences from Latin America covering only part of the RT (nucleotides 2673–3203 relative to the HXB2 clone) were also downloaded for some countries with few PR/RT sequences available (Bolivia, Suriname and French Guyana). The subtype assignment of all sequences included here was confirmed using the REGA HIV subtyping tool v.2 [8] and by performing phylogenetic analyses (see below) with HIV-1 group M subtype reference sequences. Only one sequence per subject was selected and those sequences containing frameshift mutations or with incorrect subtype assignment were removed. This resulted in a final data set of 7,654 subtype B *pol* sequences isolated from 18 Latin American countries between 1989 and 2011. These sequences were aligned with subtype B *pol* (PR/RT) sequences from the US (n = 165), France (n = 135), the Caribbean (n = 279) and Panama (n = 37), representative of the $B_{PANDEMIC}$ and the B_{CAR} clades as described previously [5, 7]. Sequences were aligned using the Clustal W program [9] and all sites associated with major antiretroviral drug resistance in PR (30, 32, 46, 47, 48, 50, 54, 76, 82, 84, 88 and 90) and RT (41, 65, 67, 69, 70, 74, 100, 101, 103, 106, 115, 138, 151, 181, 184, 188, 190, 210, 215, 219 and 230) were excluded. All alignments are available from the authors upon request.

Phylogenetic analysis. Maximum Likelihood (ML) phylogenetic trees were inferred under the GTR+I+Γ nucleotide substitution model selected using the jModeltest program [10]. The ML trees were reconstructed with the PhyML program [11] using an online web server

[12]. 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*) [13] based on the Shimodaira-Hasegawa-like procedure. The ML trees were visualized using the FigTree v1.4.0 program [14].

Analysis of the spatiotemporal dispersion pattern. The evolutionary rate, the age of the most recent common ancestor (T_{MRCA}) and the spatial diffusion pattern of non-pandemic HIV-1 subtype B clades circulating in South America were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 [15-16] with BEAGLE to improve run-time [17]. Analyses were performed using the GTR+I+Γ_4 nucleotide substitution model, a relaxed uncorrelated lognormal molecular clock model [18], and a Bayesian Skyline coalescent tree prior [19]. The mean evolutionary rates previously estimated for the subtype B *pol* gene (2.0-2.5 x 10^{-3} subst./site/year) [7, 20-22] were incorporated as an informative prior interval. Migration events throughout the phylogenetic history and the most relevant migration pathways were reconstructed using a reversible discrete phylogeography model and the Bayesian stochastic search variable selection (BSSVS) approach [23], with a CTMC rate reference prior [24]. Three MCMC chains were run for 500 x 10^6 generations and then combined using LogCombiner v1.8. Convergence and uncertainty of parameter estimates were assessed by calculating the Effective Sample Size (ESS) and 95% Highest Probability Density (HPD) values, respectively, after excluding the initial 10% of each run with Tracer v1.6 [25]. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.8 and visualized with FigTree v1.4.0. Migratory events were summarized using the cross-platform SPREAD application [26].

RESULTS

Detection of HIV-1 B_{CAR} clades in the majority of Latin American countries.

In order to estimate the relative prevalence of pandemic ($B_{PANDEMIC}$) and non-pandemic (B_{CAR}) subtype B lineages in Latin America, *pol* (PR/RT) sequences from different Latin American countries were divided into six subsets: Central America (n = 688), Mexico (n = 1,677), Argentina (n = 1,548), Brazil-I (n = 1,329), Brazil-II (n = 1,329), and other South American countries (n = 909). A seventh subset containing shorter subtype B *pol* (RT) sequences from some Latin American countries poorly represented in the PR/RT dataset (Bolivia = 45, French Guyana = 108, Suriname = 21) was also constructed. Each of the seven Latin American subsets was combined with a reference subtype B dataset selected from a previous study [5] containing 500 sequences representative of the $B_{PANDEMIC}$ (US/France = 300) and the B_{CAR} (Caribbean = 200) clades (Table S1). The ML analyses of all PR/RT (Fig. 1A and Fig. S1) and RT (Fig. 1B) subsets confirmed the complete segregation of the $B_{PANDEMIC}$ reference sequences in a highly supported (aLRT > 0.90) monophyletic clade nested within basal B_{CAR} reference sequences. The ML analyses also confirmed the circulation of B_{CAR} sequences in most Latin American countries, although with highly variable prevalence (Fig. 2 and Table S2). The B_{CAR} sequences reach a high prevalence (40-50%) in French Guyana and Suriname; low prevalence (1-10%) in Brazil, Colombia, Ecuador, Mexico, Panama and Venezuela; and very low prevalence (<1%) in Argentina, El Salvador, Honduras and Peru. We found no evidence of circulation of B_{CAR} clades in Bolivia and Chile. The number of PR/RT or RT sequences from Belize, Costa Rica, Guatemala, Guyana, Nicaragua, Paraguay and Uruguay was too small (n < 10) to allow any conclusion about the relative prevalence of different subtype B clades circulating in those Latin America countries.

Spatiotemporal dispersal pattern of the HIV-1 B_{CAR} clades in Latin America.

To reconstruct the origin and spatiotemporal dynamics of non-pandemic subtype B Latin American lineages, the HIV-1 B_{CAR} PR/RT sequences with known sampling date from Latin America here identified (n = 103) were combined with B_{CAR} PR/RT sequences from the most widely sampled (n > 10) Caribbean islands (Dominican Republic [n = 123], Jamaica [n = 73], Trinidad and Tobago [n = 50], and Haiti [n = 12]) and from Panama (n = 37), previously identified [5, 7]. The B_{CAR} sequences were further aligned with subtype D PR/RT sequences (n = 10) from the Democratic Republic of Congo (DRC) that was pointed as the most probable source of subtype B strain introduced in the Americas [4]. HIV-1 subtypes B and D sequences were classified into 14 discrete geographic locations (Table S3) and subjected to Bayesian phylogeographic analysis.

The mean estimated evolutionary rate of the HIV-1 B_{CAR}/D *pol* dataset was 2.1 x 10^{-3} substitutions/site per year (95% HPD 2.0 x 10^{-3} – 2.2 x 10^{-3} substitutions/site per year), whereas the corresponding median coefficient of rate variation was 0.31 (95% HPD: 0.27 – 0.35), supporting the selection of a relaxed molecular clock model. The root location of the HIV-1 subtype B ancestor was most probably placed in the island of Hispaniola (Dominican Republic/Haiti) (posterior state probability [*PSP*] = 0.92) (Fig. 3), consistent with previous findings [4-5]. The median estimated T_{MRCA} of subtypes B/D (1956), subtype D (1968) and subtype B (1968) were also very similar to that previously obtained using different *pol* and *env* datasets [4-5] (Table 1). The close match of major spatiotemporal calibration points across different studies validates the time-scale inferred from this analysis and indicates that the overall phylogeographic reconstruction was quite robust to the inclusion of new B_{CAR} sequences from Latin America.

193 After the introduction of HIV-1 subtype B into Hispaniola around the middle 1960s, non-
194 pandemic B_{CAR} lineages were independently disseminated to other countries from the
195 Caribbean and Latin America from the early 1970s onwards. Some of those viral
196 migrations seeded secondary outbreaks that resulted in the origin of several country-
197 specific B_{CAR} subclades including those previously identified in Trinidad and Tobago
198 ($B_{CAR\text{-}TT}$) [4-5], Jamaica ($B_{CAR\text{-}JM\text{-}I}$) [5] and Panama ($B_{CAR\text{-}PA\text{-}I}$, $B_{CAR\text{-}PA\text{-}II}$ and $B_{CAR\text{-}PA\text{-}III}$)
199 [7], and others here identified in Argentina ($B_{CAR\text{-}AR}$), Brazil ($B_{CAR\text{-}BR\text{-}I}$, $B_{CAR\text{-}BR\text{-}II}$ and
200 $B_{CAR\text{-}BR\text{-}III}$), Guyana ($B_{CAR\text{-}GY}$), Mexico ($B_{CAR\text{-}MX\text{-}I}$, and $B_{CAR\text{-}MX\text{-}II}$) and Venezuela ($B_{CAR\text{-}VE}$)
201 (Fig. 3). The non-pandemic clades $B_{CAR\text{-}TT}$, $B_{CAR\text{-}JM\text{-}I}$ and $B_{CAR\text{-}BR\text{-}I}$ seem to have originated
202 around the early 1970s, whereas most of the remaining country-specific B_{CAR} clades
203 probably arose between the late 1970s and the middle 1980s (Fig. 3 and Table 1).
204 Reconstruction of viral migrations across time suggests that Hispaniola was the major hub
205 of dissemination of non-pandemic subtype B clades in the region and further identified a
206 few secondary hubs in the Caribbean (Trinidad and Tobago) and South America (Brazil
207 and Guyana) (Figs. 4A and 4B). The $B_{CAR\text{-}TT}$ clade was independently disseminated from
208 Trinidad and Tobago to other Caribbean islands and to several South American countries
209 including Brazil, Guyana (originating the $B_{CAR\text{-}GY}$ clade), Suriname and Venezuela. The
210 $B_{CAR\text{-}GY}$ clade was disseminated from Guyana to Suriname and the $B_{CAR\text{-}BR\text{-}I}$ clade was
211 disseminated from Brazil to Argentina at multiple times (originating the $B_{CAR\text{-}AR}$ clade).
212 The Bayes factor tests for significant nonzero rates supports epidemiological linkage
213 between Hispaniola and most other Caribbean and Latin American countries included in the
214 study (with exception of Argentina and Guyana) as well as between Trinidad and Tobago
215 and Jamaica/Guyana/Brazil, between Brazil and Argentina, and between Guyana and
216 Suriname (Figs. 4C and 4D and Table S4).

DISCUSSION

The HIV-1 subtype B virus was probably originally introduced into Haiti seeded by the epidemic from the DRC around the middle 1960s [4]. After a short period of local expansion within the island of Hispaniola (shared by Haiti and the Dominican Republic), the virus seems to have moved out on several independent occasions. The introduction of the virus into the US around the late 1960s explosively amplified the number of new cases of HIV-1 subtype B infection and originates a $B_{PANDEMIC}$ strain that was disseminated across the world [4]. Other secondary outbreaks simultaneously emerged in the Caribbean [5] and Latin America [6-7] as the result of short-distance disseminations of non-pandemic B_{CAR} strains out of Hispaniola. This study demonstrates that B_{CAR} strains reached nearly all countries in Latin America, although their prevalence is usually much lower than that estimated for the $B_{PANDEMIC}$ clade (Fig. 2). The only exceptions in the region were Suriname, French Guyana and probably Guyana, where both $B_{PANDEMIC}$ and B_{CAR} clades seem to circulate at roughly similar prevalence.

Our results indicate that Haiti and Dominican Republic, that together are home to about 75% of people living with HIV in the Caribbean [27], were probably the major sources of B_{CAR} lineages disseminated into the region. Non-pandemic B_{CAR} strains started to spread from Hispaniola in the beginning of the 1970s and would have reached Trinidad and Tobago, Jamaica, Brazil, Colombia, Ecuador, El Salvador, Honduras, Mexico, Panama, Suriname and Venezuela in the following years. Trinidad and Tobago can be viewed as a secondary hub, seeding tertiary B_{CAR} outbreaks in short-distanced countries such as Jamaica, Venezuela, Guyana and Brazil. Jamaica, by contrast, seems to have played a minor role in the regional dispersion of B_{CAR} strains. We also identified short-distance spreading of B_{CAR} lineages from Brazil to Argentina and from Guyana to Suriname,

241 indicating that some South American countries also acted as secondary hubs of
242 dissemination of non-pandemic subtype B lineages in the region.
243 Although Dominican Republic, Haiti and Trinidad and Tobago were pointed as the most
244 important sources of B_{CAR} lineages disseminated to Latin America, we can not ruled out the
245 possible role of other Caribbean islands with high prevalence of B_{CAR} strains such as
246 Martinique, Guadeloupe and other Lesser Antilles [5] not included in our phylogeographic
247 analysis because the very low numbers ($n < 10$) of PR/RT sequences available. This
248 geographical sampling bias may have resulted in an overestimation of the role of
249 Dominican Republic, Haiti and Trinidad and Tobago as source of B_{CAR} lineages in the
250 region. The use of more geographically balanced HIV-1 subtype B Caribbean datasets will
251 be of paramount importance to obtain more precise estimates of the contribution of each
252 Caribbean island in the regional dissemination of non-pandemic subtype B strains.
253 Several country-specific B_{CAR} clades were detected in Argentina, Brazil, Guyana, Mexico,
254 Panama and Venezuela, suggesting that despite their overall low prevalence, non-pandemic
255 subtype B lineages have been disseminated locally in several Latin American countries.
256 Estimation of the T_{MRCA} of those country-specific B_{CAR} clades further suggests that B_{CAR}
257 lineages started to be disseminated from the Caribbean into Latin America between the
258 early 1970s and the early 1980s. This time-scale coincides with the global dissemination of
259 the $B_{PANDEMIC}$ clade from the US [4] and with the estimated origin of several $B_{PANDEMIC}$
260 lineages in Latin America [7, 28]. Although the $B_{PANDEMIC}$ and the B_{CAR} clades probably
261 arrived at the same time in Latin America, the $B_{PANDEMIC}$ strain was able to ignite much
262 larger outbreaks and infected a much larger number of individuals than any B_{CAR} strain in
263 most of the countries analyzed.

The different epidemic outcomes of the $B_{PANDEMIC}$ and B_{CAR} lineages in Latin America could be related to virological and/or sociological factors. Notably, the highest HIV prevalence rates (>1%) in Latin America and the Caribbean were detected among countries with a high proportion (≥ 50%) of B_{CAR} clades like Haiti, Bahamas, Guyana, Jamaica, and Trinidad and Tobago [5], thus arguing against the hypothesis of a low epidemic potential of B_{CAR} lineages. Transmission route is clearly an important factor shaping the HIV dissemination dynamics and major differences in the epidemic outcome of distinct subtype B clades may have appeared as a consequence of differences in the underlying transmission networks. We suggest that in most Latin American countries the $B_{PANDEMIC}$ strain was introduced and initially disseminated within highly connected networks of MSM and injecting drug users, whereas the B_{CAR} clades were mainly disseminated through heterosexual networks with lower rates of partner exchanges, which may explain the more successful dissemination of the $B_{PANDEMIC}$ lineage.

The remarkably successful dissemination of B_{CAR} clades in some northern countries of South America including French Guyana, Suriname and Guyana, probably reflects the high mobility of people between these countries and the Caribbean islands [29]. This is facilitated not only by the geographical proximity of those South American countries to the Caribbean islands, but also by cultural, linguistic and socioeconomic ties. Suriname and Guyana are members of the Caribbean Common Market (CARICOM), an organization of 15 Caribbean nations and dependencies that also includes Bahamas, Belize, Haiti, Jamaica, Trinidad and Tobago and several other Lesser Antilles islands. The CARICOM not only promotes economic integration, but also facilitates the free movement of individuals for tourism or labor among countries. It notes that a significant proportion (10%) of immigrants residing in Trinidad and Tobago are from Guyana [29], which may explain the

288 epidemiological link observed between non-pandemic B_{CAR-TT} and B_{CAR-GY} clades
289 circulating in Trinidad and Tobago and Guyana, respectively.
290 The higher frequency of B_{CAR} clades in Colombia, Panama and Venezuela (4-9% of
291 subtype B infections) when compared to other Latin American countries (<2% of subtype B
292 infections) also probably reflects a more frequent population mobility as a consequence of
293 greater geographical proximity and historical links. It is interesting to note that the first
294 reported Panamanian AIDS case was a Haitian woman diagnosed in September 1984 [30],
295 which supports a longstanding presence of viruses of Caribbean origin in Panama. This
296 country is also an important commercial hub due to the presence of the Panama Canal that
297 promotes transit of people and goods. Junqueira *et al* (2011) previously noted that a boom
298 in oil production in Venezuela attracted immigrants from several countries in the region
299 between 1970 and 1980, including people from Trinidad and Tobago and the Dominican
300 Republic, which may have promoted the introduction of B_{CAR} strains into Venezuela during
301 that time. Furthermore, Colombia and Venezuela has been pointed out as the most
302 important source countries in South America for tourists and labor migrants (including
303 female sex workers) to many Caribbean islands (particularly in the Netherlands Antilles)
304 [29].
305 In summary, this study demonstrates that several non-pandemic HIV-1 B_{CAR} strains have
306 been disseminated from the Caribbean into Latin America since the early 1970s. The B_{CAR}
307 strains reached nearly all countries from Latin America here analyzed and in some of them
308 were spread locally, establishing secondary outbreaks. Despite the early and widespread
309 dissemination of B_{CAR} strains in the continent, HIV-1 subtype B epidemics in most Latin
310 American countries were mainly driven by the $B_{PANDEMIC}$ clade that accounts for most (>
311 90%) of current HIV-1 subtype B infections in the region. The only exceptions were

312 Suriname, French Guyana and probably Guyana, where both $B_{PANDEMIC}$ and B_{CAR} clades
313 seem to circulate at roughly similar prevalence as observed in many Caribbean islands.
314 Intra-regional population mobility combined with chance founder events in populations
315 with high rates of partner exchange were probably the major forces driving the actual
316 distribution of the different subtype B strains in the Americas.
317
318 **ACKNOWLEDGMENTS**
319 We wish to thank Dr. Vera Bongertz for critical review of the manuscript.
320 Contributions: The study was conceived and designed by G. Bello. Data acquisition was
321 performed by M. Cabello. All authors contributed to the data analysis and final version of
322 the paper.
323
324 **REFERENCES**
325 1. UNAIDS. Report on the global AIDS epidemic. Accessed at
326 http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/201
327 3/gr2013/UNAIDS_Global_Report_2013_en.pdf) 2013.
328 2. UNAIDS. Global HIV/AIDS response. Progress Report 2011. Accessed at:
329 http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/
330 2011/20111130_UA_Report_en.pdf) 2011.
331 3. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular
332 epidemiology of HIV-1 during 2000-2007. *Aids* 2011,**25**:679-689.
333 4. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The
334 emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A*
335 2007,**104**:18566-18570.
336 5. Cabello M, Mendoza Y, Bello G. Spatiotemporal Dynamics of Dissemination of
337 Non-Pandemic HIV-1 Subtype B Clades in the Caribbean Region. *PLoS ONE*
338 2014,**9**:e106045.
339 6. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, Ashton-Prolla P,
340 *et al.* Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS*
341 *ONE* 2011,**6**:e27489.
342 7. Mendoza Y, Martinez AA, Castillo Mewa J, Gonzalez C, Garcia-Morales C, Avila-
343 Rios S, *et al.* Human Immunodeficiency Virus Type 1 (HIV-1) Subtype B Epidemic
344 in Panama Is Mainly Driven by Dissemination of Country-Specific Clades. *PLoS*
345 *ONE* 2014,**9**:e95360.

8. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C*, et al.* An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005,**21**:3797-3800.
9. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997,**25**:4876-4882.
10. Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 2008,**25**:1253-1256.
11. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010,**59**:307-321.
12. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005,**33**:W557-559.
13. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 2006,**55**:539-552.
14. Rambaut A. FigTree v1.4: Tree Figure Drawing Tool. *Available from* http://tree.bio.ed.ac.uk/software/figtree/ 2009.
15. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 2002,**161**:1307-1320.
16. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007,**7**:214.
17. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. *Bioinformatics* 2009,**25**:1370-1376.
18. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006,**4**:e88.
19. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005,**22**:1185-1192.
20. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci U S A* 2005,**102**:4425-4429.
21. Zehender G, Ebranati E, Lai A, Santoro MM, Alteri C, Giuliani M*, et al.* Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. *J Acquir Immune Defic Syndr* 2010,**55**:156-160.
22. Chen JH, Wong KH, Chan KC, To SW, Chen Z, Yam WC. Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS ONE* 2011,**6**:e25286.
23. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009,**5**:e1000520.
24. Ferreira MAR, M.A. S. Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics* 2008,**26**:355-368.
25. Rambaut A, Drummond A. Tracer v1.6. *Available from* http://tree.bio.ed.ac.uk/software/tracer/ 2007.
26. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 2011,**27**:2910-2912.

27. UNAIDS. AIDS INFO DATABASE. *Accessed at http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/* 2013.
28. Murillo W, Veras N, Prosperi M, de Rivera IL, Paz-Bailey G, Morales-Miranda S*, et al.* A single early introduction of HIV-1 subtype B into Central America accounts for most current cases. *J Virol* 2013,**87**:7463-7470.
29. Borland R, Faas L, Marshall D, McLean R, Schroen M, Smit M*, et al. HIV/AIDS and mobile populations in the Caribbean: a baseline assessment*: International Organization for Migration; 2004.
30. de Ycaza MM, Rios V, Miranda E, Narvaez E, Sanchez G. [Acquired immune deficiency syndrome. First confirmed case in Panama]. *Rev Med Panama* 1985,**10**:66-73.

Table 1. Bayesian time-scale estimates of MRCA of HIV-1 subtypes B and D and major B_{CAR} clades from Latin America and the Caribbean.

Clade	T_{MRCA} Current study	T_{MRCA} Cabello *et al* (2014)	T_{MRCA} Gilbert *et al* (2007)
Subtypes B/D	1956 (1946-1963)	1952 (1943-1960)	1954 (1946-1961)
Subtype D	1968 (1961-1973)	1965 (1958-1971)	1966 (1961-1971)
Subtype B	1968 (1963-1972)	1964 (1959-1969)	1966 (1962-1970)
B_{CAR-TT}	1973 (1969-1976)	1969 (1966-1973)	1973 (1970-1976)
$B_{CAR-JM-I}$	1973 (1969-1976)	1971 (1967-1975)	-
$B_{CAR-JM-II}$	1982 (1977-1987)	-	-
$B_{CAR-BR-I}$	1973 (1971-1977)	-	-
$B_{CAR-BR-II}$	1979 (1975-1983)	-	-
$B_{CAR-BR-III}$	1983 (1977-1987)	-	-
$B_{CAR-MX-I}$	1980 (1974-1986)	-	-
$B_{CAR-MX-II}$	1981 (1976-1987)	-	-
$B_{CAR-PA-I}$	1977 (1973-1981)	-	-
$B_{CAR-PA-II}$	1980 (1976-1984)	-	-
$B_{CAR-PA-III}$	1989 (1983-1995)	-	-
B_{CAR-AR}	1979 (1975-1982)	-	-
B_{CAR-VE}	1978 (1974-1983)	-	-
B_{CAR-GY}	1980 (1977-1984)	-	-

Table S1. HIV-1 subtype B *pol* (PR/RT and RT) sequences from Latin America, the Caribbean, US and France used for ML phylogenetic analyses.

Region	Country	*N* (PR/RT)	*N* (RT)	Sampling time
Latin America	Argentina	1,548	-	1998-2009
	Bolivia	9	45	1999-2009
	Brazil	2,658	-	1990-2010
	Chile	118	-	2002-2007
	Colombia	58	-	2000-2010
	Ecuador	44	-	1989-2011
	El Salvador	170	-	2008-2010
	French Guyana	-	108	2000-2002
	Honduras	507	-	2001-2009
	Peru	249	-	2003-2010
	Mexico	1,677	-	2004-2010
	Suriname	5	21	2000
	Venezuela	407	-	2004-2011
	Others [a]	30	-	1999-2009
Caribbean	Dominican Republic	61	-	2005-2010
	Haiti	8	-	2004-2005
	Jamaica	62	-	2005-2010
	Trinidad and Tobago	48	-	2000-2003
	Others [b]	21	-	2000-2004
North America	US	165	-	1997-2009
Europe	France	135	-	1985-2008

[a] Belize (n = 9), Costa Rica (n = 2), Guyana (n = 6), Paraguay (n = 5), and Uruguay (n = 8).

[b] Antigua and Barbuda (n = 4), Bahamas (n = 5), Dominica (n = 1), Grenada (n = 2), Montserrat (n = 1), Saint Lucia (n = 4) and Saint Vincent and the Grenadines (n = 4).

Table S2. Classification of HIV-1 subtype B subtype *pol* (PR/RT and RT) sequences from different Latin American countries.

Region	Country	*N*	$B_{PANDEMIC}$	B_{CAR}
South America	Argentina	1,548	1,534 (99.1%)	14 (0.9%)
	Bolivia	54	54 (100%)	0 (0%)
	Brazil	2,658	2,614 (98.3%)	44 (1.7%)
	Chile	118	118 (100%)	0 (0%)
	Colombia	58	53 (91.4%)	5 (8.6%)
	Ecuador	44	43 (97.7%)	1 (2.3%)
	French Guyana	108	48 (44.5%)	60 (55.5%)
	Guyana	6	0 (0%)	6 (100%)
	Paraguay	5	5 (100%)	0 (0%)
	Peru	249	247 (99.2%)	2 (0.8%)
	Suriname	26	12 (46.1%)	14 (53.8%)
	Uruguay	8	8 (100%)	0 (0%)
	Venezuela	407	391 (96.1%)	16 (3.9%)
Central America	Belize	9	9 (100%)	0 (0%)
	Costa Rica	2	2 (100%)	0 (0%)
	El Salvador	170	169 (99.4%)	1 (0.6%)
	Honduras	507	506 (99.8%)	1 (0.2%)
	Panama*	761	719 (99.5%)	42 (0.5%)
North America	Mexico	1,677	1,659 (98.9%)	18 (1.1%)

* Estimated from a previous study [7].

Table S3. HIV-1 B_{CAR} *pol* (PR/RT) sequences from Latin America and the Caribbean used for Bayesian phylogeographic analysis.

Region	Country	Location	*N*	Sampling date
South America	Argentina	AR	12	2001-2007
	Brazil	BR	39	1997-2005
	Colombia	CO	3	2001-2002
	Ecuador	EC/PE	1	2005
	Peru	EC/PE	2	2009-2010
	Guyana	GY	6	2000
	Suriname	SR	5	2000
	Venezuela	VE	15	2004-2009
Central America	El Salvador	HN/SV	1	2002
	Honduras	HN/SV	1	2008
	Panama [a]	PA	37	2004-2013
North America	Mexico	MX	18	2006-2010
Caribbean	Dominican Republic [a]	DO/HT	123	2003-2011
	Haiti [a]	DO/HT	12	2004-2005
	Jamaica [a]	JM	73	2005-2010
	Trinidad and Tobago [a]	TT	50	2000-2003
Central Africa	DRC	CD	10	1983-2007

[a] Identified in previous studies [5, 7].

Table S4. Bayes factor (BF) rates of epidemiological links between locations for dispersal of non-pandemic B_{CAR} lineages in the Latin America.

Locations	BF*
HIS-VE	45,724
HIS-PA	45,724
HIS-MX	45,724
HIS-JM	45,724
HIS-BR	45,724
HIS-PE/EC	2,685
TT-GY	891
BR-AR	410
GY-SR	234
HIS-SR	190
HIS-CO	127
HIS-HN/SV	95
HIS-TT	66
TT-JM	35
TT-BR	12
DRC-HIS	8
Others	<3

AR: Argentina; BR: Brazil; CO: Colombia; DRC: Democratic Republic of Congo; GY: Guayana; HIS: Hispaniola; HN/SV: Honduras/El Salvador; JM: Jamaica; MX: Mexico; PA: Panamá; PE/EC: Peru/Ecuador; SR: Suriname; TT: Trinidad and Tobago; VE: Venezuela. *BF > 100 indicates decisive support, 30 ≤ BF ≤ 100 indicates very strong support, 10 ≤ BF ≤ 30 indicates strong support, and 6 ≤ BF ≤ 10 indicates substantial support for migration between locations.

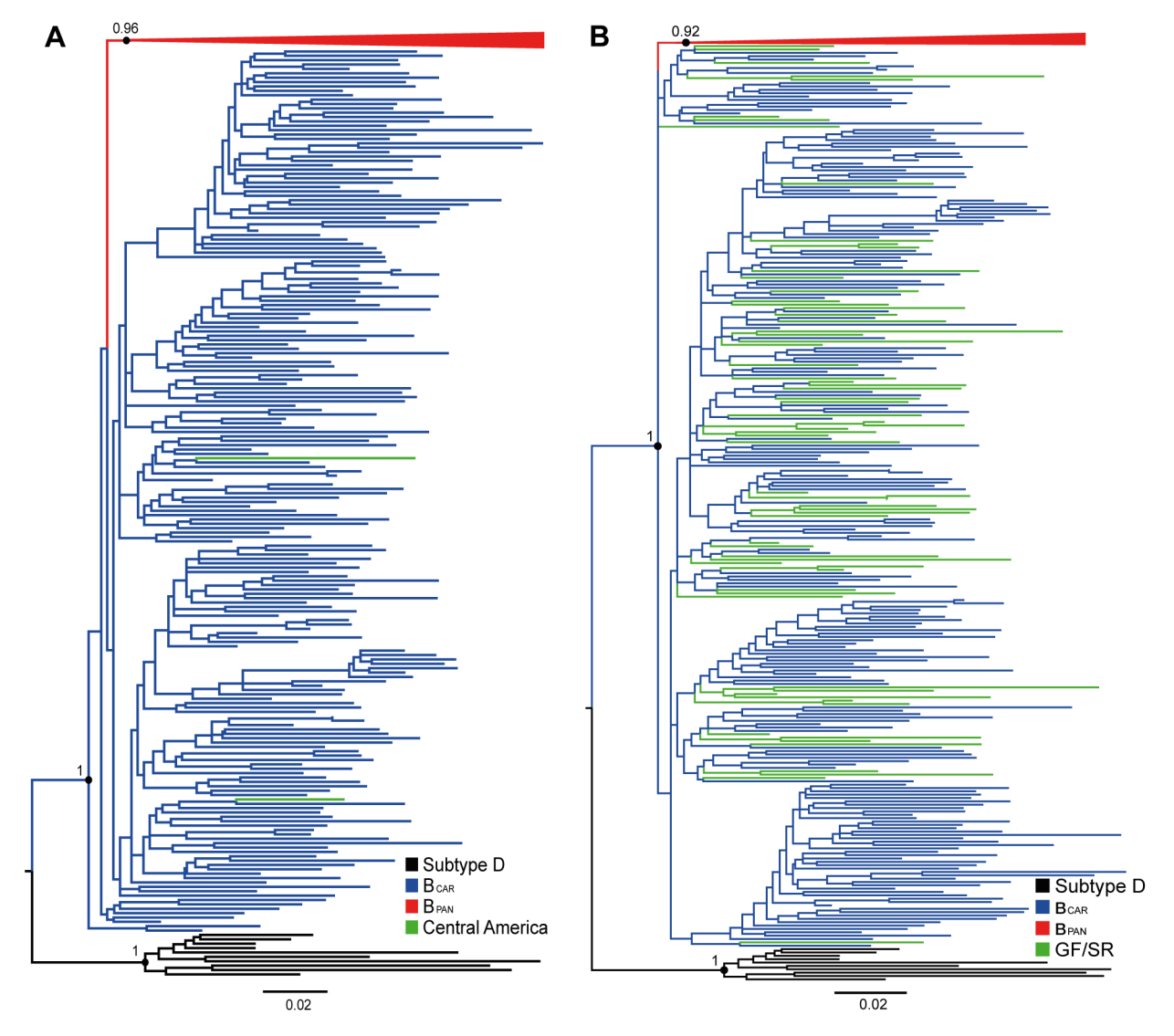
FIGURE LEGENDS

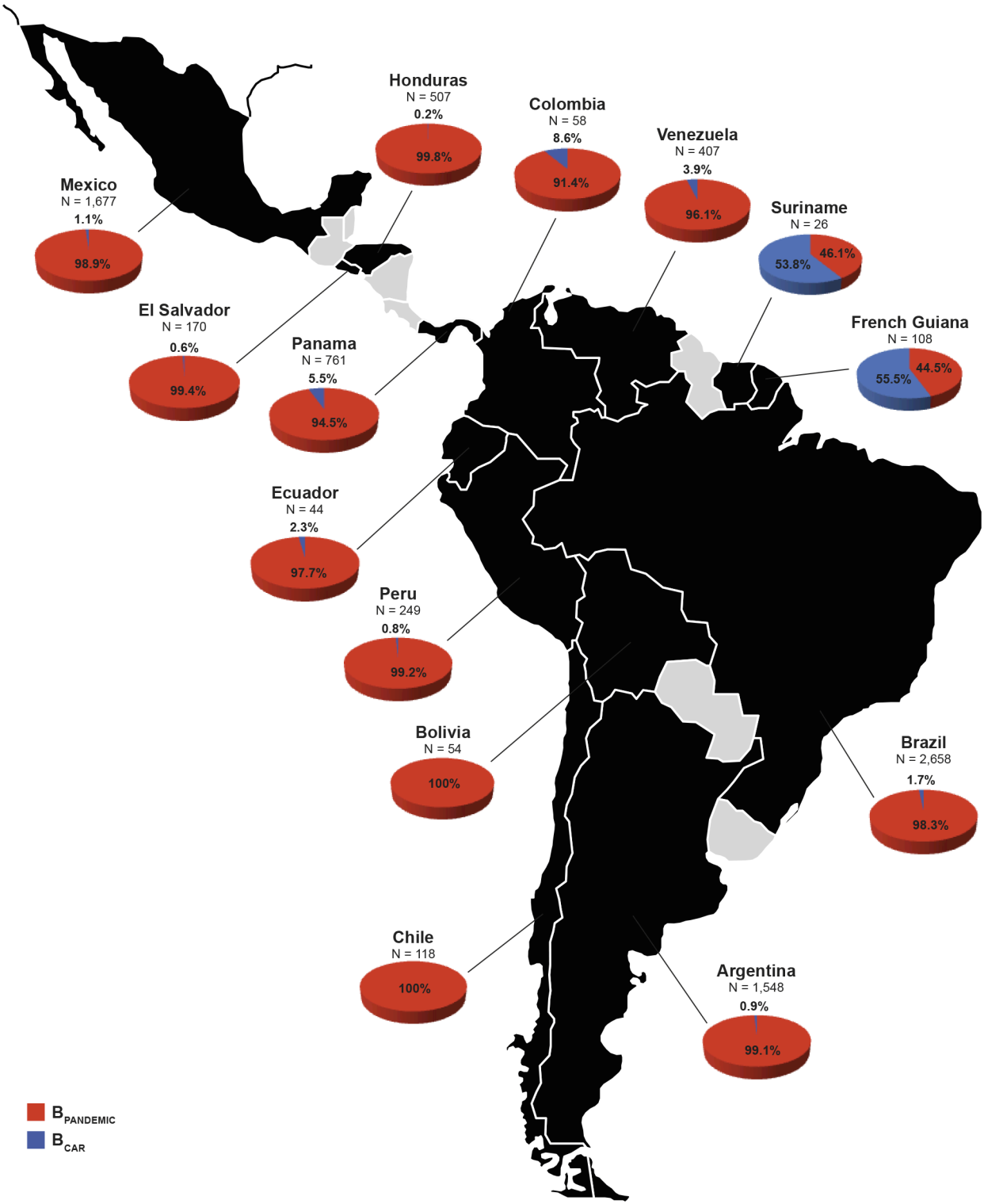
Figure 1. ML phylogenetic tree of A) HIV-1 subtype B *pol* PR/RT sequences (~1,000 nt) circulating in Central America (n = 688) and representative sequences of the $B_{PANDEMIC}$ (US = 165, France = 135) and the B_{CAR} (Caribbean = 200) clades; B) HIV-1 subtype B pol RT (~600 nt) sequences from Bolivia (n = 45), French Guyana (n = 108), Suriname (n = 21) and the representative sequences of the $B_{PANDEMIC}$ and the B_{CAR} clades. Branches are colored according to the geographic origin/clade classification of each sequence as indicated at the legend (bottom right). The $B_{PANDEMIC}$ clade was collapsed for visual clarity. The aLRT support values are indicated at key nodes. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

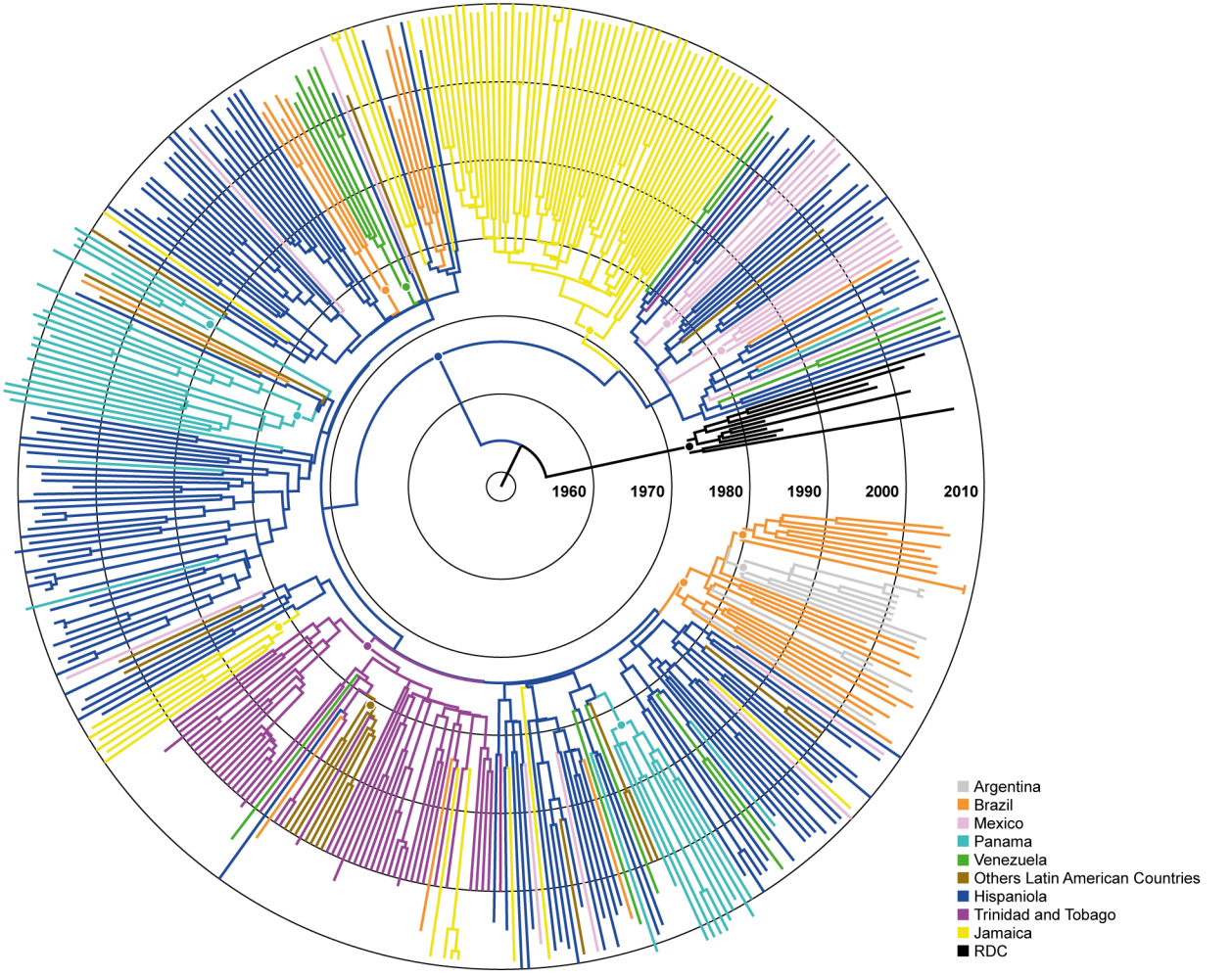
Figure 2. Estimated proportion of B_{CAR} and $B_{PANDEMIC}$ clades among HIV-1 subtype B infected individuals from different Latin American countries according to the ML analyses. The total number of sequences analyzed in each locality is indicated. Proportions in Panama were estimated in a previous study [7]. Proportions in Latin American countries poorly sampled ($n < 10$) were not estimated.

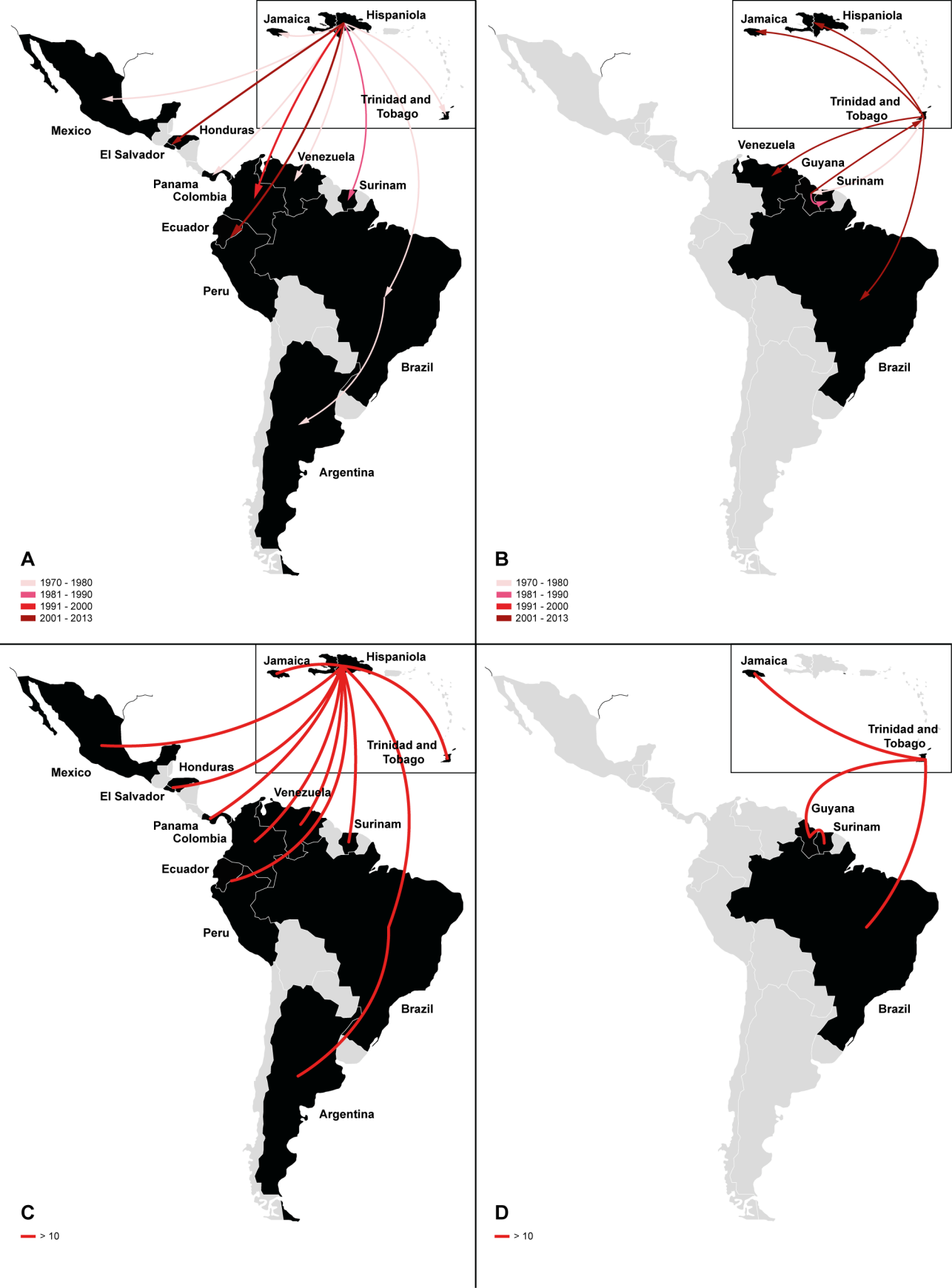
Figure 3. Time-scaled Bayesian MCMC tree of *pol* PR/RT sequences of HIV-1 B_{CAR} lineages from Latin America and the Caribbean, and subtype D reference sequences from the Democratic Republic of Congo (DRC). Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (bottom right). Colored circles indicate the positions of nodes corresponding to the most recent common ancestors of major country-specific clades (clade size ≥ 4). Branch lengths are depicted in

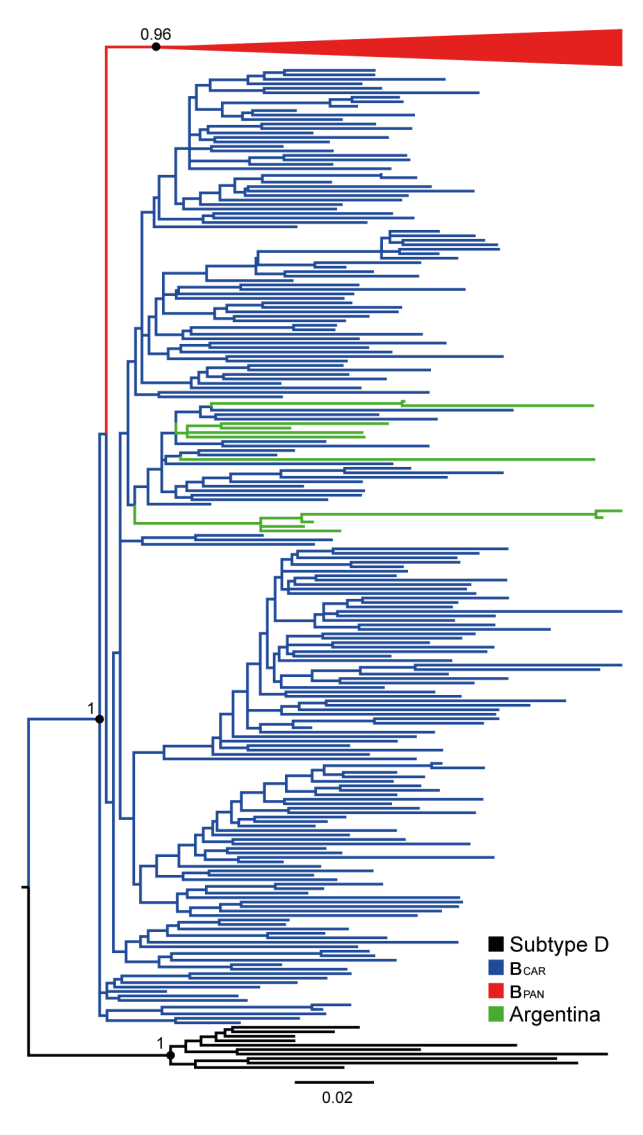
458 units of time (years). The tree was automatically rooted under the assumption of a relaxed
459 molecular clock.
460
461 **Figure 4.** Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 B_{CAR} clades
462 in Latin America. A and B) Viral migration events occurred between 1970 and 2013 are
463 indicated. Lines between locations represent branches in the Bayesian MCC tree along
464 which location transitions occurred. The line's color informs the estimated years of the viral
465 migrations and only the earliest transitions between each location pair were represented. C
466 and D) Most significant epidemiological links of the dissemination process of B_{CAR} clades.
467 Only epidemiological links supported by Bayes factor rates > 3 are displayed. Viral
468 migrations and most significant epidemiological links connecting the Hispaniola (A and C)
469 and Trinidad and Tobago (B and D) with Latin American countries were separated in
470 independent panels only for visual clarity.
471
472 **Figure S1.** ML phylogenetic tree of HIV-1 subtype B *pol* PR/RT sequences (~1,000 nt)
473 circulating in: A) Argentina (*n* = 1,548), B) Brazil (*n* = 2,658), C) other South American
474 countries (*n* = 909), and D) Mexico (*n* = 1,677) combined with representative sequences of
475 the $B_{PANDEMIC}$ (US = 165, France = 135) and the B_{CAR} (Caribbean = 200) clades. Branches
476 are colored according to the geographic origin/clade classification of each sequence as
477 indicated at the legend (bottom right). The $B_{PANDEMIC}$ clade was collapsed for visual clarity.
478 The aLRT support values are indicated at key nodes. Trees were rooted using HIV-1
479 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the
480 bottom indicating nucleotide substitutions per site.
481

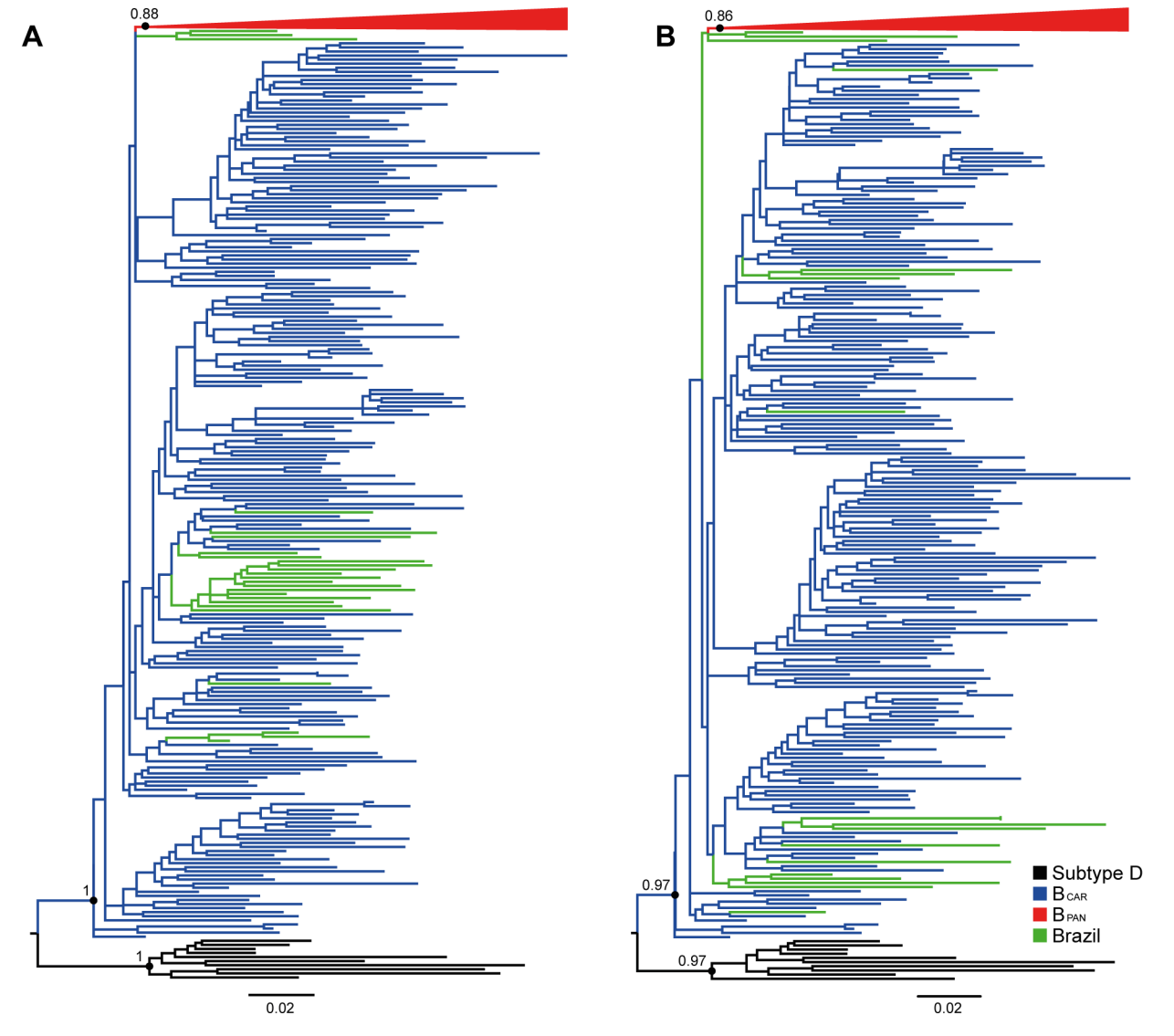

A
0.96
1
1
Subtype D
B CAR
B PAN
Central America
0.02
B
0.92
1
1
Subtype D
B CAR
B PAN
GF/SR
0.02

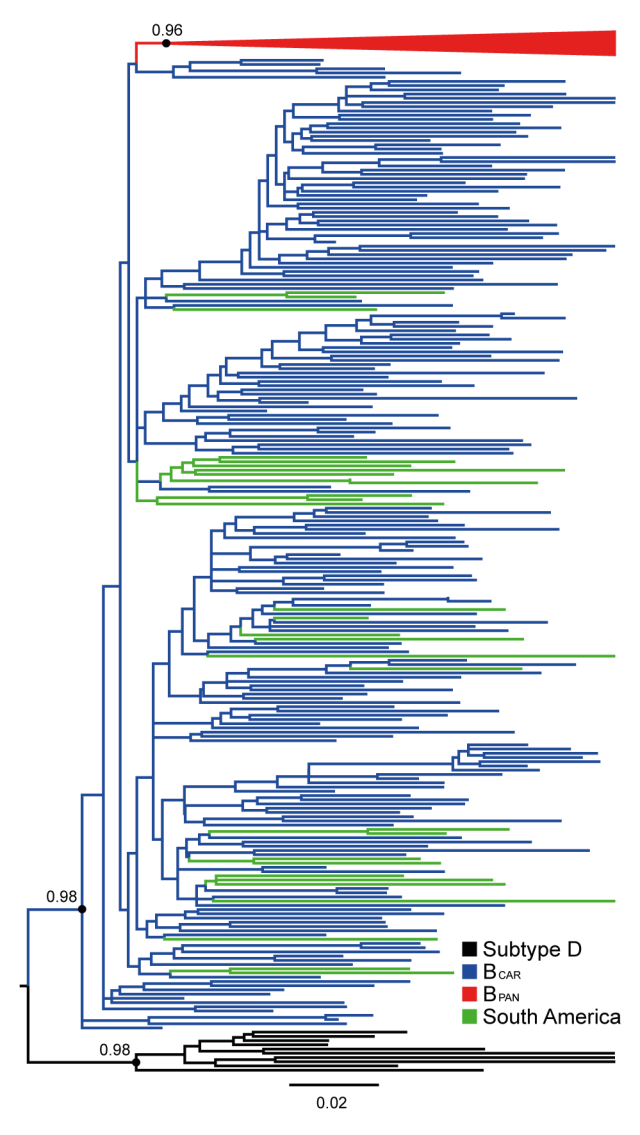

Honduras
N = 507
0.2%
99.8%
Colombia
N = 58
8.6%
91.4%
Venezuela
N = 407
3.9%
96.1%
Suriname
N = 26
46.1%
53.8%
Mexico
N = 1,677
1.1%
98.9%
El Salvador
N = 170
0.6%
99.4%
Panama
N = 761
5.5%
94.5%
French Guiana
N = 108
44.5%
55.5%
Ecuador
N = 44
2.3%
97.7%
Peru
N = 249
0.8%
99.2%
Bolivia
N = 54
100%
Brazil
N = 2,658
1.7%
98.3%
Chile
N = 118
100%
Argentina
N = 1,548
0.9%
99.1%
B PANDEMIC
B CAR

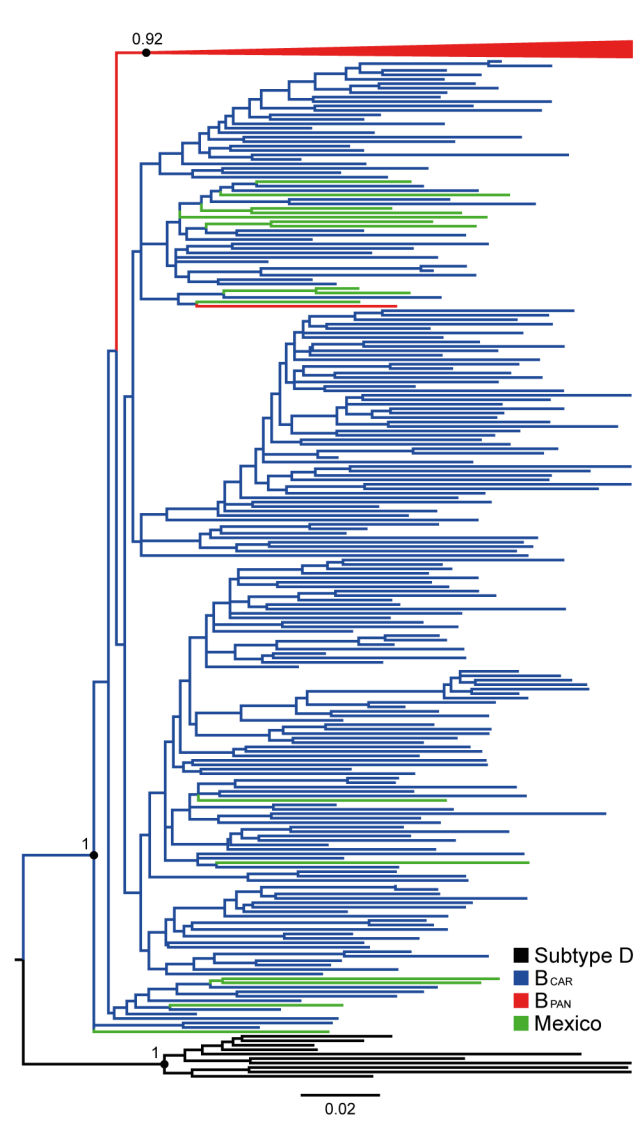

1960
1970
1980
1990
2000
2010
Argentina
Brazil
Mexico
Panama
Venezuela
Others Latin American Countries
Hispaniola
Trinidad and Tobago
Jamaica
RDC


Jamaica
Hispaniola
Trinidad and Tobago
Mexico
Honduras
El Salvador
Venezuela
Panama
Colombia
Surinam
Ecuador
Peru
Brazil
Argentina
A
1970 - 1980
1981 - 1990
1991 - 2000
2001 - 2013
Jamaica
Hispaniola
Trinidad and Tobago
Venezuela
Guyana
Surinam
Brazil
B
1970 - 1980
1981 - 1990
1991 - 2000
2001 - 2013
Jamaica
Hispaniola
Trinidad and Tobago
Mexico
Honduras
El Salvador
Venezuela
Panama
Colombia
Surinam
Ecuador
Peru
Brazil
Argentina
C
> 10
Jamaica
Trinidad and Tobago
Guyana
Surinam
Brazil
D
> 10


0.96
1
1
Subtype D
BCAR
BPAN
Argentina
0.02


A
0.88
1
1
0.02
B
0.86
0.97
0.97
0.02
Subtype D
B CAR
B PAN
Brazil


0.96
0.98
0.98
Subtype D
B CAR
B PAN
South America
0.02


0.92
1
1
Subtype D
B CAR
B PAN
Mexico
0.02