

SCIENTIFIC REPORTS



OPEN

Multiple introductions and onward transmission of non-pandemic HIV-1 subtype B strains in North America and Europe

Marina Cabello¹, Hector Romero² & Gonzalo Bello¹

Received: 01 June 2016

Accepted: 06 September 2016

Published: 22 September 2016

Most HIV-1 subtype B infections in North America and Europe seem to have resulted from the expansion of a single pandemic lineage (B_{PANDEMIC}) disseminated from the United States (US). Some non-pandemic subtype B strains of Caribbean origin (B_{CAR}) may have also reached North America and Europe, but their epidemiological relevance in those regions remains largely unknown. Here we analyze a total of 20,045 HIV-1 subtype B *pol* sequences from the US, Canada, and Europe, to estimate the prevalence and to reconstruct the spatiotemporal dynamics of dissemination of HIV-1 B_{CAR} strains in those regions. We find that B_{CAR} strains were probably disseminated from the Caribbean into North America and Europe at multiple times since the early 1970s onwards. The B_{CAR} strains reached the US, Canada and at least 16 different European countries, where they account for a very low fraction (<5%) of subtype B infections, with exception of the Czech Republic (7.7%). We also find evidence of the onward transmission of B_{CAR} clades in the US, Canada, the Czech Republic, Germany, Italy, Spain and the UK, as well as short-distance spreading of B_{CAR} lineages between neighboring European countries from Central and Western Europe, and long-distance dissemination between the US and Europe.

By the end of 2012, the total number of people living with the Human Immunodeficiency Virus Type 1 (HIV-1) reached an estimated 1.3 million [980.000–1.900.000] in North America and 860.000 [800.000–930.000] in Western/Central Europe¹. Unprotected sex between men who have sex with men (MSM) is the main driver of HIV transmission in the US, Canada and Europe, followed by heterosexual contact and injecting drug use (IDU)^{2–4}. The proportion of newly diagnosed HIV cases attributed to MSM increased in the US, Canada and Europe over the last 10 years, while the corresponding proportions attributed to heterosexual contact and IDU remained stable or decreased in the same period^{2–4}.

Subtype B dominates the HIV-1 epidemic in North America and in Western and Central Europe⁵; although each country has a unique HIV molecular pattern and an increasing prevalence of non-B subtypes have been observed in the United States (US)⁶, Canada⁷, and Europe^{8–11} in the last decade. The HIV-1 subtype B epidemic in North America and Europe probably started via a single viral introduction from Haiti into the US around the late 1960s, followed by rapid viral dissemination within the US and from the US to other American countries and Europe, leading to the origin of a pandemic subtype B clade (B_{PANDEMIC})¹². Networks of MSM and IDUs probably played a crucial role in the early spread of the B_{PANDEMIC} lineage within North America and from North America to Europe^{13–15}. The HIV-1 subtype B epidemic in most European countries originated from multiple sources^{16,17}, revealing the importance of human mobility across international borders in the spread of this clade. Current subtype B transmissions in Europe, however, seem to be predominantly occurring between patients from the same country¹⁸.

The study of Gilbert *et al.*¹² also pointed to the existence of non-pandemic subtype B lineages that, in sharp contrast to the B_{PANDEMIC} clade, remained mostly restricted to the Caribbean region (B_{CAR} clades). More recent studies conducted by our group demonstrate that several non-pandemic B_{CAR} lineages have been disseminated out of Haiti since the early 1970s and reach a high prevalence (40–95%) in several countries throughout the

¹Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil. ²Laboratorio de Organizacion y Evolucion del Genoma, Departamento de Ecologia y Evolucion, Facultad de Ciencias-CURE, Universidad de la Republica, Montevideo, Uruguay. Correspondence and requests for materials should be addressed to G.B (email: gbello@ioc.fiocruz.br)

Caribbean region including Haiti, Dominican Republic, Jamaica, Trinidad and Tobago, other Lesser Antilles, Suriname and French Guiana^{19–21}. The non-pandemic B_{CAR} lineages did not remain restricted to the Caribbean region, but also reached most countries from Latin America²⁰ and despite the overall low prevalence (<10%), established secondary outbreaks of small size in Argentina, Brazil, Mexico, Panama and Venezuela^{20,21}. North America and Europe have maintained a high level of population movement with several Caribbean countries^{22–24}, but the epidemiological relevance of non-pandemic B_{CAR} strains in those regions has not been explored.

The objective of this study was to estimate the prevalence of the HIV-1 B_{PANDEMIC} and B_{CAR} clades in North America and Europe and to reconstruct the spatiotemporal dynamics of dissemination of non-pandemic subtype B lineages detected in those regions. For this, we used a comprehensive dataset of HIV-1 subtype B *pol* sequences ($n = 20,045$) isolated from the US, Canada, and 26 European countries between 1982 and 2012. North American and European subtype B sequences were combined with reference sequences representative of the B_{PANDEMIC} and the B_{CAR} clades and then analyzed using Maximum Likelihood and Bayesian phylogeographic approaches.

Materials and Methods

HIV-1 subtype B *pol* sequence dataset. We downloaded all HIV-1 subtype B *pol* sequences from North America ($n = 9,670$) and Europe ($n = 10,885$) 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 July 2014. Only one sequence per subject was selected and those sequences containing frameshift mutations (1.4%) or with incorrect subtype B assignment (1.1%) according to the REGA HIV subtyping tool v.2²⁵ were removed. This resulted in a final data set of 20,045 subtype B *pol* sequences isolated from North America (Canada and the US) and from Central (Austria, Germany, Poland, Slovakia and Switzerland), Northern (Denmark, Finland, Norway and Sweden), Southern (Italy, Portugal and Spain), Southeastern (Albania, Cyprus, Greece, Montenegro, Romania, Serbia and Slovenia) and Western (Belgium, Ireland, Luxembourg, Netherlands, France and the United Kingdom) Europe, between 1982 and 2012 (Table S1).

Subtype B Clade assignment. HIV-1 subtype B *pol* sequences from North America and Europe were aligned with subtype B *pol* sequences representative of the B_{PANDEMIC} and the B_{CAR} clades described previously^{19,21}, using the Clustal W program²⁶. Codons associated with major antiretroviral drug resistance in PR ($n = 12$) and RT ($n = 21$) were excluded, resulting in a final alignment of 909 nucleotides. Sequences were initially classified as B_{CAR} or B_{PANDEMIC} by using an evolutionary placement algorithm (EPA)²⁷ available in RAxML²⁸ for the rapid assignment of query sequences to edges of a reference phylogenetic tree under a maximum-likelihood (ML) model. Sequences classified within the B_{CAR} clade were again combined with the B_{PANDEMIC} and the B_{CAR} reference sequences and subjected to new rounds of ML phylogenetic analysis with the PhyML program²⁹ using an online web server³⁰, for confirmation of clade assignment. Trees were inferred under the GTR+I+ Γ nucleotide substitution model, selected using the jModeltest program³¹, and the SPR branch-swapping algorithm of heuristic tree search. The reliability of tree topology was estimated with the approximate likelihood-ratio test (*aLRT*)³² based on the Shimodaira-Hasegawa (SH)-like procedure. Clusters were classified as medium (SH-*aLRT* = 0.85–0.94) and well (SH-*aLRT* \geq 0.95) supported, considering that the false positive error rate should not exceed 10% and 1% respectively³³. Potential epidemiologic North American and European B_{CAR} clades were defined as those strongly supported monophyletic clusters exclusively composed by sequences from those geographic regions. The ML trees were visualized using the FigTree v1.4.0 program³⁴.

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 North America, Europe and the Caribbean were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8^{35,36} with BEAGLE to improve run-time³⁷. Analyses were performed using the GTR+I+ Γ_4 nucleotide substitution model, a relaxed uncorrelated lognormal molecular clock model³⁸, and a Bayesian Skyline coalescent tree prior³⁹. The regression analysis of root-to-tip genetic distance against sampling time performed by using the program TempEst⁴⁰, revealed that the HIV-1 subtype B dataset here compiled does not contain sufficient temporal signal for reliable time-scale estimation (X-intercept [T_{MRCA}] = 1908). Thus, we specified a uniform prior distribution on the substitution rate ($2.0\text{--}3.0 \times 10^{-3}$ subst./site/year) that encompass mean values previously estimated for the subtype B *pol* gene^{16,21,41,42}. 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) approach⁴³, with a CTMC rate reference prior⁴⁴. Discrete locations were assigned according to the sampling country of the sequences, except for European sequences that were grouped according to the sampling region (Central, Northern, Southern, Southeastern and Western, see Table S1) to reduce the complexity of the migration matrix. Three MCMC chains were run for 5×10^8 generations and then combined using LogCombiner v1.8 while excluding the initial states (10–20%) of each run, needed to reach convergence. Convergence and uncertainty of parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively with Tracer v1.6⁴⁵. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.8 and visualized with FigTree v1.4.0. Assuming that the posterior probability (*PP*) support closely reflected the probability for a split to be correct (under the true evolutionary model and correct priors distributions)⁴⁶, clades were defined as moderate ($PP = 0.85\text{--}0.94$) or strongly ($PP \geq 0.95$) supported. The cross-platform SPREAD application⁴⁷ was used to summarize the migratory events and to identify statistically well-supported migration routes (those with Bayes factor [BF] > 3).

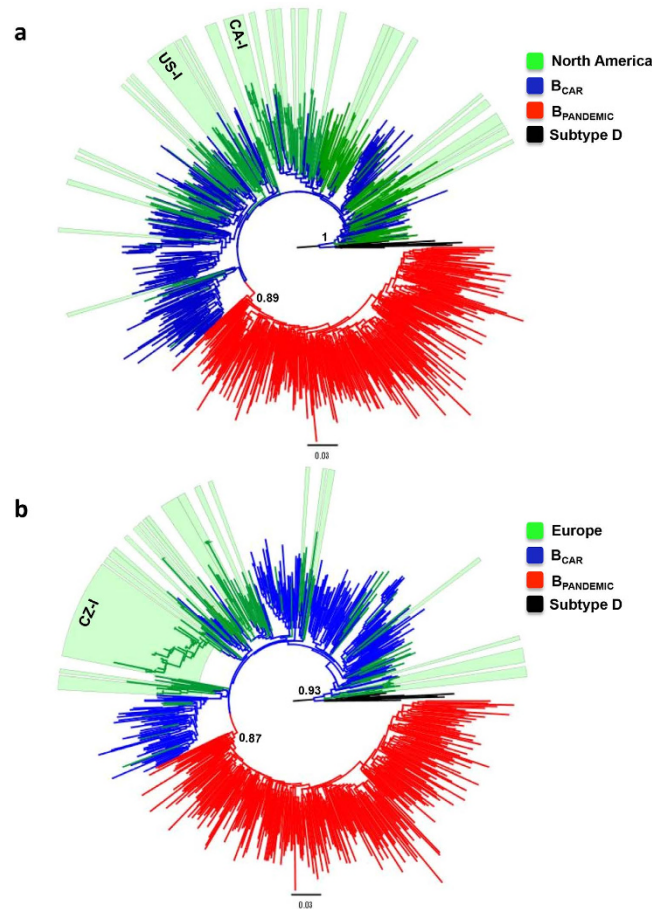


Figure 1. ML phylogenetic trees of HIV-1 B_{CAR} *pol* sequences identified in North America (a) and Europe (b). North American ($n = 274$) and European ($n = 189$) B_{CAR} sequences were combined with representative sequences of the $B_{PANDEMIC}$ (US = 165, France = 135) and the B_{CAR} (Caribbean = 200) clades. Branches are colored according to the geographic origin/clade classification of each sequence as indicated at the legend (upper right). Colored boxes show the positions of strongly supported ($SH-aLRT \geq 0.95$) North American and European B_{CAR} clades. Major ($n \geq 10$) B_{CAR} clades detected in North America (US-I and CA-I) and Europe (CZ-I) are indicated. $SH-aLRT$ supports for subtype B and $B_{PANDEMIC}$ clades are shown. 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.

Results

Detection of HIV-1 B_{CAR} strains in North America and Europe. A total of 20,045 HIV-1 subtype B *pol* sequences from different countries from North America and Europe were combined with a set of $B_{PANDEMIC}$ and B_{CAR} reference sequences¹⁹ and classified by using two sequential ML-based frameworks. This procedure classified a total of 274 (2.9%) and 189 (1.8%) HIV-1 subtype B *pol* sequences from North America and Europe within the B_{CAR} clades, respectively. The final ML phylogenetic trees clearly showed that the B_{CAR} sequences from North America and Europe were dispersed among the B_{CAR} reference sequences of Caribbean origin, while all $B_{PANDEMIC}$ reference sequences branched in a medium supported ($SH-aLRT = 0.85-0.90$) clade nested within basal B_{CAR} reference sequences (Fig. 1). These analyses confirmed the circulation of B_{CAR} sequences in Canada, the US and 16 out of 26 European countries here analyzed (Fig. 2). Most European countries with detection of B_{CAR} strains encompass larger number of sequences ($n > 250$) than European countries with no evidence of circulation of B_{CAR} strains ($n < 100$) (Table S1). The B_{CAR} strains reached a low prevalence (<5%) among subtype B-infected individuals from all North American and European countries analyzed, with the only exception of the Czech Republic (CZ) where the B_{CAR} strains represent 7.7% of the sequences included (Fig. 2 and Table S1).

The ML phylogenetic analyses also revealed that many B_{CAR} strains from North America and Europe appear as sporadic lineages intermixed among Caribbean sequences, whereas other branched in strongly supported ($aLRT \geq 0.95$) clades exclusively composed by sequences from those regions (Fig. 1). Most North American and European B_{CAR} clades were of small size ($n < 10$) and comprise sequences from a single country (country-specific clades) (Fig. S1 and Table S2). A few country-specific clades of large size ($n = 11-48$), however, were detected in Canada, the US, and the CZ (Fig. 1). We also detected a few international clades that encompass sequences from at least two different countries (Canada/US, Germany/Netherlands, CZ/Portugal/UK, Belgium/CZ/Norway, Germany/Netherlands/Switzerland/UK and CZ/Germany/Spain/Switzerland) (Fig. S1). In most countries

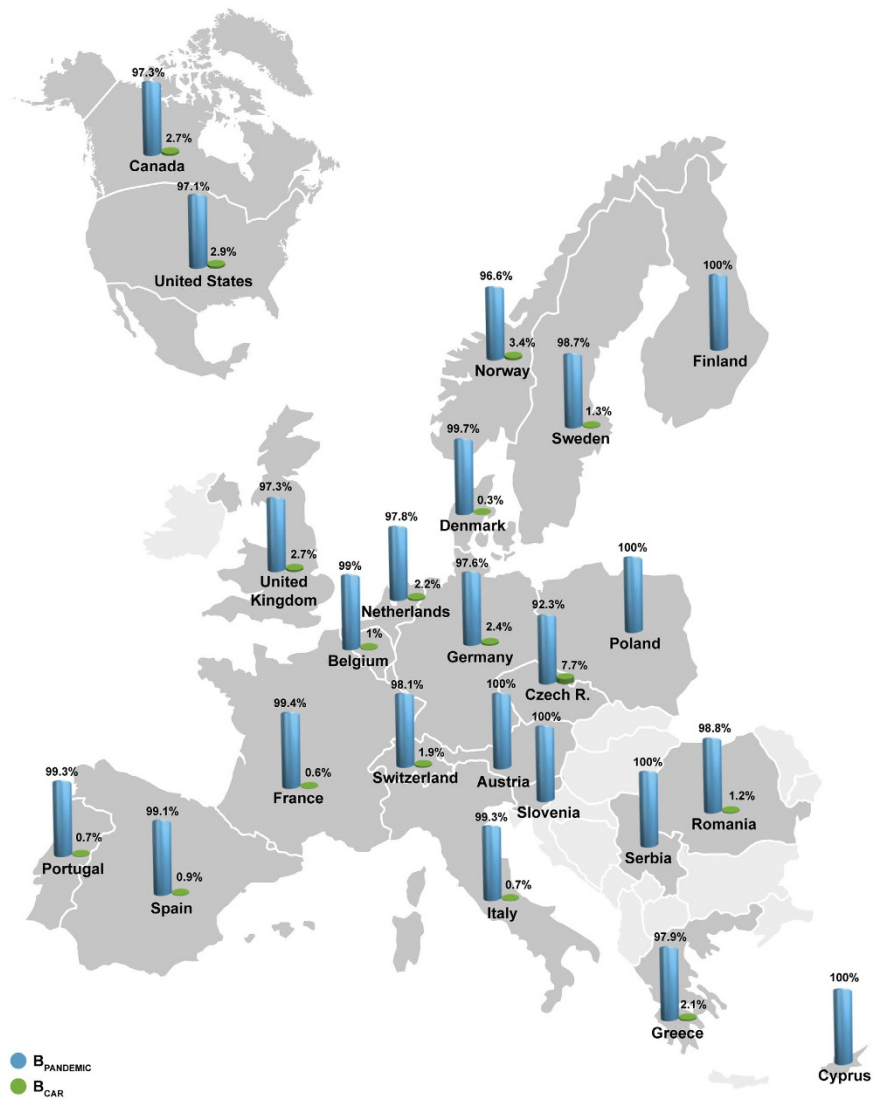


Figure 2. Estimated proportion of B_{CAR} and $B_{PANDEMIC}$ clades among HIV-1 subtype B infected individuals from different North American and European countries. Maps were created with Adobe Illustrator CC from templates obtained from d-maps.com (North America: http://d-maps.com/pays.php?num_pay=119&lang=es; and Europe: http://d-maps.com/pays.php?num_pay=192&lang=es).

analyzed the B_{CAR} sequences were mostly classified as sporadic lineages (57–67%), with exception of Canada, Germany and the CZ, where most B_{CAR} sequences (60–97%) were classified within country-specific/international clades (Table S2).

Origin of HIV-1 B_{CAR} strains introduced in North America and Europe. HIV-1 B_{CAR} *pol* sequences with known sampling date from North America ($n = 216$) and Europe ($n = 126$) here identified, were next combined with B_{CAR} *pol* sequences from the most widely sampled ($n > 10$) Caribbean islands ($n = 258$)^{19,21} and with subtype D *pol* sequences (the closest relative to subtype B) from the Democratic Republic of Congo (the most probable epicenter of ancestral subtype B)¹² ($n = 10$) that served as outgroup (Table S3). HIV-1 sequences were classified into 10 discrete geographic locations and subjected to Bayesian phylogeographic analyses. When all B_{CAR} sequences were combined in a single dataset, the MCMC chains fail to converge and many parameters showed low (< 100) ESS values despite very long runs (1×10^9 generations). Furthermore, the most probable ancestral root location for the subtype B epidemic was traced to the US (posterior state probability [PSP] > 0.90), which is inconsistent with the seminal work of Gilbert and colleagues (2007) that clearly traced the origin of American subtype B epidemic to Haiti. Sampling bias can confuse the accurate estimation of the spatial root of an epidemic⁴⁸ and our convenience sampling was clearly biased to the US ($n = 164$) (Table S3). Moreover, the B_{CAR} US sequences were widely scattered among B_{CAR} sequences of Hispaniola (probably due to many independent viral introductions into the US), so the B_{CAR} genetic diversity in both locations was comparable and the simple overrepresentation of the US may bias the ancestral root location to this country.

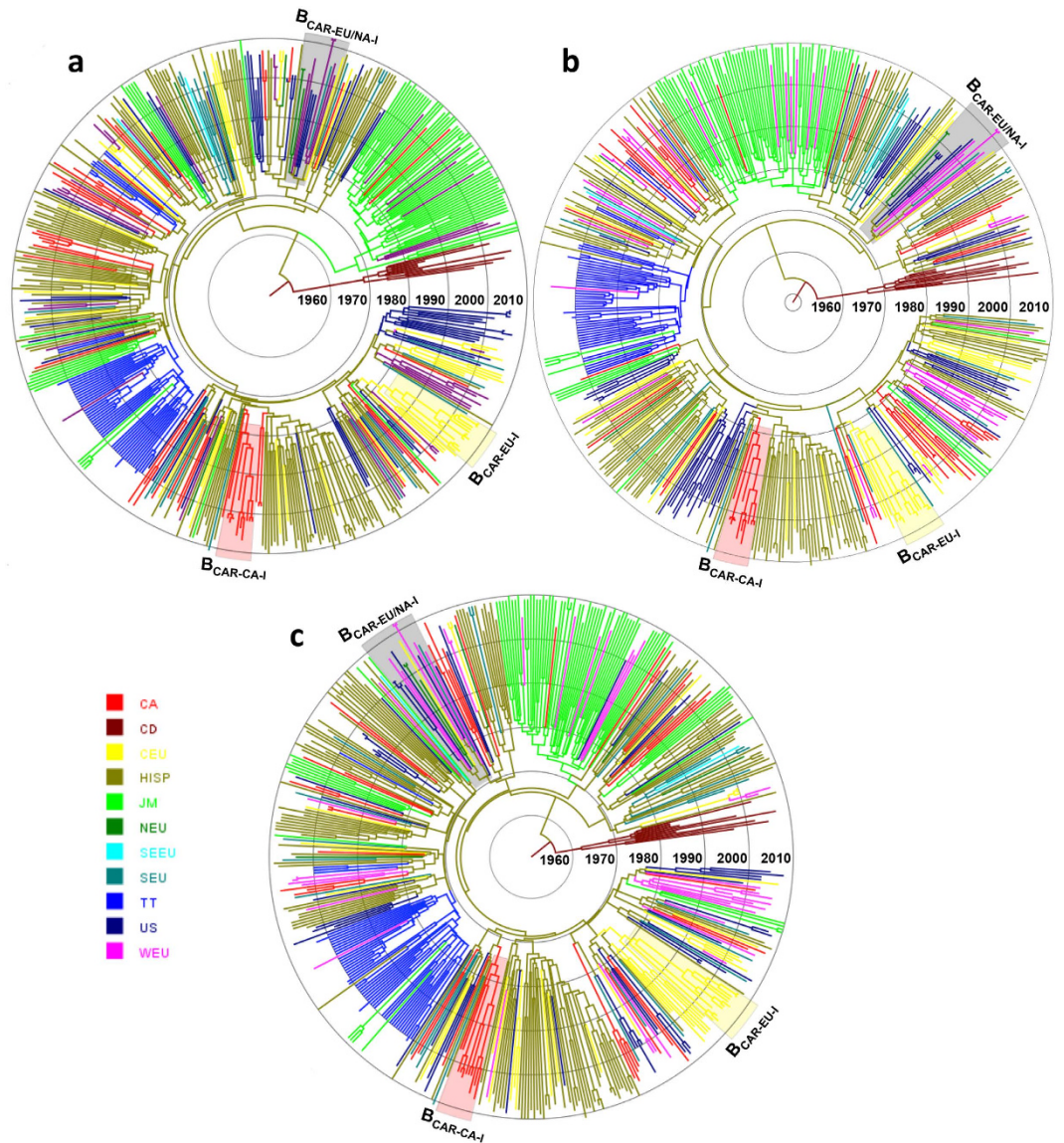


Figure 3. Time-scaled Bayesian MCMC tree of HIV-1 B_{CAR} *pol* sequences from North America and Europe. North American ($n = 216$) and European ($n = 126$) B_{CAR} sequences with known sampling date were combined with B_{CAR} sequences from the Caribbean ($n = 258$), $B_{PANDEMIC}$ sequences ($n = 50$) the US and France and subtype D sequences ($n = 10$) from the DRC. Branches of trees reconstructed from subsets 1 (a), 2 (b), and 3 (c) are colored according to the most probable location state (see Table S3) of their descendent nodes as indicated in the legend (bottom left). Colored boxes indicate the positions of major ($n \geq 10$) B_{CAR} clades detected outside the Caribbean. Branch lengths are depicted in units of time (years). The trees were automatically rooted under the assumption of a relaxed molecular clock. CA: Canada, CD: Democratic Republic of Congo, CEU: Central Europe, HISP: Hispaniola, JM: Jamaica, NEU: Northern Europe, SEEU: Southeastern Europe, SEU: Southern Europe, TT: Trinidad and Tobago, US: United States of America, WEU: Western Europe.

In order to reduce the impact of sampling bias on the accuracy of phylogeographic reconstructions, sequences from the US were subdivided into three subsets ($n = 54$ – 55) at random, except for those sequences belonging to the small country-specific clades ($n \leq 5$) that were placed in the same subset (Table S4) in order to recover their T_{MRCA} . Each subset was then combined with sequences from the other locations and independently analyzed. The overall pattern of spatial and temporal dissemination of B_{CAR} strains reconstructed from all subsets was roughly similar (Fig. 3). All three analyses pointed to the island of Hispaniola as the most probable root location of the HIV-1 subtype B ancestor ($PSP = 0.76$ – 0.99), and traced the median T_{MRCA} of subtype B at around the middle 1960s (Table 1), fully consistent with previous estimates^{12,49,50}. Phylogeographic reconstructions also suggest that Hispaniola was the major hub of international dissemination of B_{CAR} strains, sending viruses to Canada, the US and all European regions at multiple times from the middle 1970s onwards (Fig. 4A). Additional viral migrations from Jamaica to Canada/US/UK and from Trinidad and Tobago to the UK were also detected (Fig. 4A). The

Clade	Countries	Subset 1 T _{MRC}	Subset 2 T _{MRC}	Subset 3 T _{MRC}
Subtypes B/D	–	1956 (1947–1964)	1956 (1948–1964)	1957 (1948–1964)
Subtype B	–	1967 (1963–1971)	1967 (1964–1970)	1968 (1964–1971)
B _{CAR-NA/EU-I}	DE/FR/IT/LU/SE/US/UK	1974 (1971–1977)	1974 (1970–1976)	1974 (1971–1977)
B _{CAR-NA/EU-II}	BE/CA/CZ/NO/US	1978 (1974–1982)	–	–
B _{CAR-NA/EU-III}	US/ES	1980 (1976–1983)	–	–
B _{CAR-EU-I}	CH/CZ/DE	1981 (1977–1985)	1981 (1977–1985)	1981 (1977–1986)
B _{CAR-EU-II}	CH/CZ/DE	1979 (1976–1983)	1979 (1975–1983)	1980 (1976–1983)
B _{CAR-EU-III}	CH/DE/UK	1992 (1986–1996)	1992 (1987–1996)	1992 (1987–1996)
B _{CAR-EU-IV}	CZ/UK	1997 (1994–2000)	1997 (1994–2000)	1997 (1994–2000)
B _{CAR-EU-V}	IT/RO	1980 (1976–1983)	1979 (1976–1983)	1980 (1976–1983)
B _{CAR-EU-VI}	DE/NL	1982 (1977–1987)	1983 (1978–1989)	1982 (1977–1987)
B _{CAR-CZ-I}	CZ	1996 (1992–1999)	1996 (1992–1999)	1996 (1992–1999)
B _{CAR-US-II}	US	–	–	1994 (1991–1997)
B _{CAR-US-III}	US	–	–	1987 (1982–1994)
B _{CAR-US-IV}	US	–	1989 (1983–1993)	–
B _{CAR-US-V}	US	2005 (2004–2006)	–	–
B _{CAR-US-VI}	US	–	–	1994 (1988–2000)
B _{CAR-US-VII}	US	–	1994 (1991–1996)	–
B _{CAR-US-VIII}	US	–	–	1985 (1982–1989)
B _{CAR-US-IX}	US	–	1991 (1985–1996)	–
B _{CAR-US-X}	US	–	–	1996 (1995–1997)
B _{CAR-US-XII}	US	–	1995 (1991–1999)	–
B _{CAR-US-XIII}	US	–	–	1995 (1994–1997)
B _{CAR-US-XIV}	US	2004 (2002–2005)	–	–
B _{CAR-US-XV}	US	–	–	2005 (2002–2006)
B _{CAR-US-XIX}	US	–	–	1982 (1978–1987)
B _{CAR-CA-I}	CA	1979 (1975–1982)	1978 (1975–1982)	1979 (1975–1983)
B _{CAR-CA-II}	CA	1996 (1988–2002)	1996 (1988–2002)	1996 (1988–2002)
B _{CAR-CA-III}	CA	1983 (1979–1987)	1983 (1979–1987)	1984 (1979–1988)
B _{CAR-CA-IV}	CA	1998 (1994–2001)	1998 (1994–2001)	1998 (1994–2001)
B _{CAR-CA-V}	CA	1987 (1981–1993)	1987 (1981–1992)	1987 (1982–1993)
B _{CAR-CA-VI}	CA	1998 (1993–2002)	1998 (1993–2002)	1998 (1993–2002)
B _{CAR-CA-VII}	CA	1998 (1997–1998)	1998 (1996–1998)	1998 (1997–1998)
B _{CAR-DE-II}	DE	1988 (1982–1993)	1987 (1982–1994)	1988 (1982–1993)
B _{CAR-DE-III}	DE	1987 (1981–1994)	1987 (1980–1993)	1988 (1982–1994)
B _{CAR-ES-I}	ES	2006 (2004–2008)	2006 (2004–2008)	2006 (2004–2008)
B _{CAR-IT-II}	IT	1989 (1983–1993)	2006 (2004–2008)	2006 (2004–2008)
B _{CAR-UK-I}	UK	1993 (1990–1996)	1993 (1990–1996)	1994 (1990–1996)
B _{CAR-UK-II}	UK	2004 (2002–2004)	2004 (2002–2004)	2004 (2002–2004)
B _{CAR-UK-III}	UK	1994 (1988–1999)	1994 (1988–1998)	1994 (1988–1999)

Table 1. Bayesian estimates of the T_{MRC} of HIV-1 B_{CAR} clades from North America and Europe.

BF tests for significant nonzero rates support epidemiological linkage between: Hispaniola/Canada, Hispaniola/US, Hispaniola/Central Europe, Hispaniola/Southern Europe, Hispaniola/Western Europe and Jamaica/Western Europe (UK) (Fig. 4B).

Onward transmission of HIV-1 B_{CAR} strains in North America and Europe. The Bayesian analysis confirms that some B_{CAR} strains seeded secondary outbreaks in North America and Europe (Fig. 3). The overall distribution of B_{CAR} sequences across sporadic lineages and country-specific/international clades was comparable to that inferred from the ML trees (Table S2). Nearly all highly supported North American and European clades previously identified also displayed a very high support in Bayesian trees ($PP = 1$) (Table S5). Bayesian analyses also recovered three medium/highly supported ($PP = 0.90–0.99$) European B_{CAR} lineages (B_{CAR-EU-I}, B_{CAR-EU-V} and B_{CAR-EU-VI}) that were observed as low/medium supported ($aLRT = 0.82–0.89$) clades in the ML trees (Table S5). Finally, the combination of North American and European sequences in the same dataset allowed the identification of three medium/highly supported ($PP = 0.88–1$) intercontinental B_{CAR} clades (Fig. 3 and Table S5). The major intercontinental clade (B_{CAR-NA/EU-I}) comprises sequences from the US and six European countries, the second one (B_{CAR-NA/EU-II}) comprises sequences from the US, Canada, Belgium, the CZ and Norway, and the last one (B_{CAR-NA/EU-III}) comprises sequences from the US and Spain.

A great proportion (40%) of North American and European B_{CAR} clades probably arose between the late 1970 s and the late 1980 s, others (46%) arose during the 1990 s, and a minor fraction (14%) arose during the 2000 s (Table 1). Most (82%) North American and European B_{CAR} clades seem to have remained confined to one single country. A few European B_{CAR} clades were disseminated across neighboring countries such as the CZ, Germany and Switzerland ($B_{CAR-EU-I}$ and $B_{CAR-EU-II}$) or Germany and Netherlands ($B_{CAR-EU-VI}$) and others were spread over longer distances, like those disseminated from Central Europe to UK ($B_{CAR-NA/EU-III}$ and $B_{CAR-NA/EU-IV}$), and from Italy to Romania ($B_{CAR-EU-V}$) (Fig. 4A). The most notable examples of long-distance dissemination, however, were the three intercontinental B_{CAR} clades that probably arose between the early 1970 s and the early 1980 s and were spread between North America and Europe.

In order to gain a better understanding of their dissemination dynamics, the three intercontinental B_{CAR} clades were combined in a single dataset with B_{CAR} strains from the Caribbean ($n = 258$), a representative sub-set of $B_{PANDEMIC}$ sequences from the US and France ($n = 50$) and subtype D sequences from the DRC ($n = 10$). The Bayesian analysis clearly reconstructed long distance migrations of clades $B_{CAR-NA/EU-I}$, $B_{CAR-NA/EU-II}$ and $B_{CAR-NA/EU-III}$ from the US to Europe at multiple times, supported by high BF rates (Fig. 5). The T_{MRCA} of the $B_{PANDEMIC}$ clade was estimated at 1969 (1965–1973), consistent with a previous study¹², of the $B_{CAR-NA/EU-I}$ clade at 1972 (1968–1976), of the $B_{CAR-NA/EU-II}$ clade at 1977 (1972–1982), and of the of the $B_{CAR-NA/EU-III}$ clade at 1978 (1973–1982). This phylogeographic analysis further suggests that clades $B_{CAR-NA/EU-I}$, $B_{CAR-NA/EU-II}$ and $B_{CAR-NA/EU-III}$ may have evolved from a single ancestor traced to the US at 1970 (1966–1974), but this result should be interpreted with caution because the support for that large intercontinental B_{CAR} clade was low ($PP = 0.59$).

Discussion

This study confirms that the HIV-1 subtype B epidemic in North America and Europe is mostly driven by the dissemination of the $B_{PANDEMIC}$ lineage, but also demonstrates several independent introductions of non-pandemic B_{CAR} strains of Caribbean origin into those regions. The dissemination of B_{CAR} strains from the Caribbean into North America and Europe should be expected considering the high human mobility between those regions. The discovery of the Americas in 15th century was the starting point for several European countries to create colonies in the Caribbean region that persisted for several centuries and many of them continue to have government ties with European countries (such as the UK, France and the Netherlands) and the US at the 21st century. The linguistic and socioeconomic links created during the colonial period certainly facilitated: (1) a large flow of labor migrants from Caribbean countries towards North America and Europe, particularly since 1970^{22–24}, and (2) a sharp increase in the number of visitors mainly from the US, Germany, the UK, France and Canada towards the Caribbean, particularly since the 1960 s when regular international airplane flights made vacations to the Caribbean more affordable^{23,51}.

Our results indicate that the island of Hispaniola was probably the major source of B_{CAR} lineages disseminated into those regions. Jamaica can be viewed as a secondary hub sending B_{CAR} strains to the US, Canada and the UK, whereas Trinidad and Tobago seems to have played a minor role in long-distance dispersion of B_{CAR} strains. It could be argued that these estimates were biased by the sampling scheme used here since, after down-sampling the US, most B_{CAR} sequences were from Hispaniola. If we considered the relative contribution of each location to the total number of B_{CAR} -infected individuals, however, our estimates are quite robust to sampling bias since all locations were overrepresented with the only exception of Hispaniola. This Caribbean island hosts about 70% of the total number of B_{CAR} -infected individuals, but only comprises 27% of B_{CAR} sequences in our subsets. Our convenience sampling, however, probably failed to recover some important dispersal routes between Caribbean countries with high prevalence of B_{CAR} lineages and strong connection with North America/Europe, like many Lesser Antilles islands, Guyana, French Guyana and Suriname^{19,20}, that were not represented in our subsets.

Some viral migration routes here recovered are fully consistent with the notion that postcolonial ties are an important driving force in the international dissemination of B_{CAR} strains. The spread of these strains out of Jamaica, for example, was traced to those countries (the US, the UK and Canada) where most Jamaicans immigrants reside and from where most tourists visiting Jamaica originate^{22–24}. Other factors apart from the historical, linguistic and socio-economic links with the Caribbean region, however, appear to be needed to explain the overall prevalence and distribution of B_{CAR} strains across different European countries. We observed a very low prevalence (<1%) of B_{CAR} strains among subtype B-infected individuals from some countries (France and Spain) that host large numbers of Caribbean immigrants and contribute many tourists to the Caribbean region, and a relative high prevalence of B_{CAR} strains (2–8%) in other countries (the CZ, Greece, Luxembourg and Norway) located outside the main migration corridors from/to the Caribbean^{22–24}.

Stochastic events may have influenced the B_{CAR} circulation in some countries like the CZ, where most (73%) B_{CAR} sequences branched in a country-specific sub-clade that was nested among basal B_{CAR} sequences from Germany and Switzerland. This supports that the high prevalence of B_{CAR} sequences observed in the CZ probably resulted from the local expansion of a single founder strain introduced from a neighboring European country. Such a founder B_{CAR} strain probably gained access to some highly interconnected network of IDU or MSM from the CZ, although we have no epidemiological information about the B_{CAR} -infected patients to test this hypothesis. Other country-specific B_{CAR} clades were also detected in the US, Canada, Germany, Italy, Spain and the UK, supporting local dissemination of non-pandemic subtype B lineages in a number of North American and European countries. In addition, we also detected short-distance spreading of B_{CAR} lineages between neighboring European countries from Central and Western Europe, and long-distance dissemination between the US and Europe.

While stochastic founder events may help to explain the dissemination of B_{CAR} strains in countries that are not part of the main migration corridors with the Caribbean, cultural factors could be a major obstacle for the onward transmission of B_{CAR} strains primarily introduced into countries strongly connected to the Caribbean. Migrant travelers might form a bridge population for HIV transmission between their country of origin and their country of residence. It has been estimated, for example, that about 6% of the Surinamese and Antillean

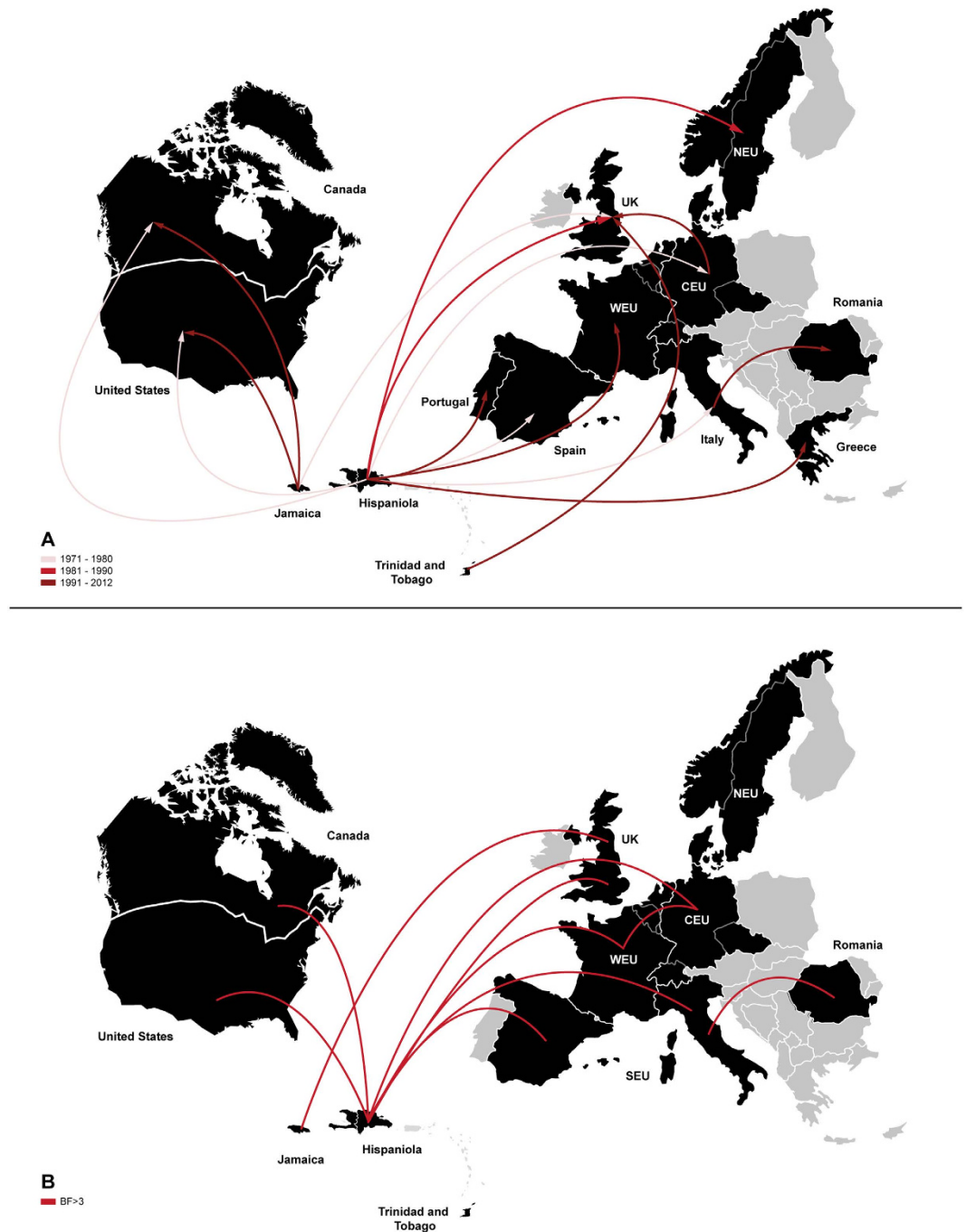


Figure 4. Spatiotemporal dynamics of dissemination of HIV-1 B_{CAR} clades from the Caribbean to North America and Europe. (a) Lines between locations represent branches in the Bayesian MCC trees along which viral migration events occurred. The line's color informs the estimated time interval of the viral migrations as indicated in the legend (bottom left). Only the earliest transitions between each location pair were represented. (b) Most significant (Bayes factor rates > 3) epidemiological links of the dissemination process of B_{CAR} clades. CEU: Central Europe, NEU: Northern Europe, WEU: Western Europe. UK: United Kingdom. Maps were created with Adobe Illustrator CC from templates obtained from d-maps.com (Caribbean: http://d-maps.com/pays.php?num_pay=118&lang=es; North America: http://d-maps.com/pays.php?num_pay=119&lang=es; and Europe: http://d-maps.com/pays.php?num_pay=192&lang=es).

migrant population act as a potential bridge population for HIV transmission in The Netherlands⁵². Molecular epidemiological data shows, however, that HIV transmission occurs mostly within migrant communities, whereas transmission between Surinamese, Antillean, and Dutch individuals living in The Netherlands are rare⁵³. Consequently, the B_{CAR} strains potentially introduced into The Netherlands from Suriname and the Netherlands Antilles should not be expected to fuel large outbreaks among indigenous Dutch population.

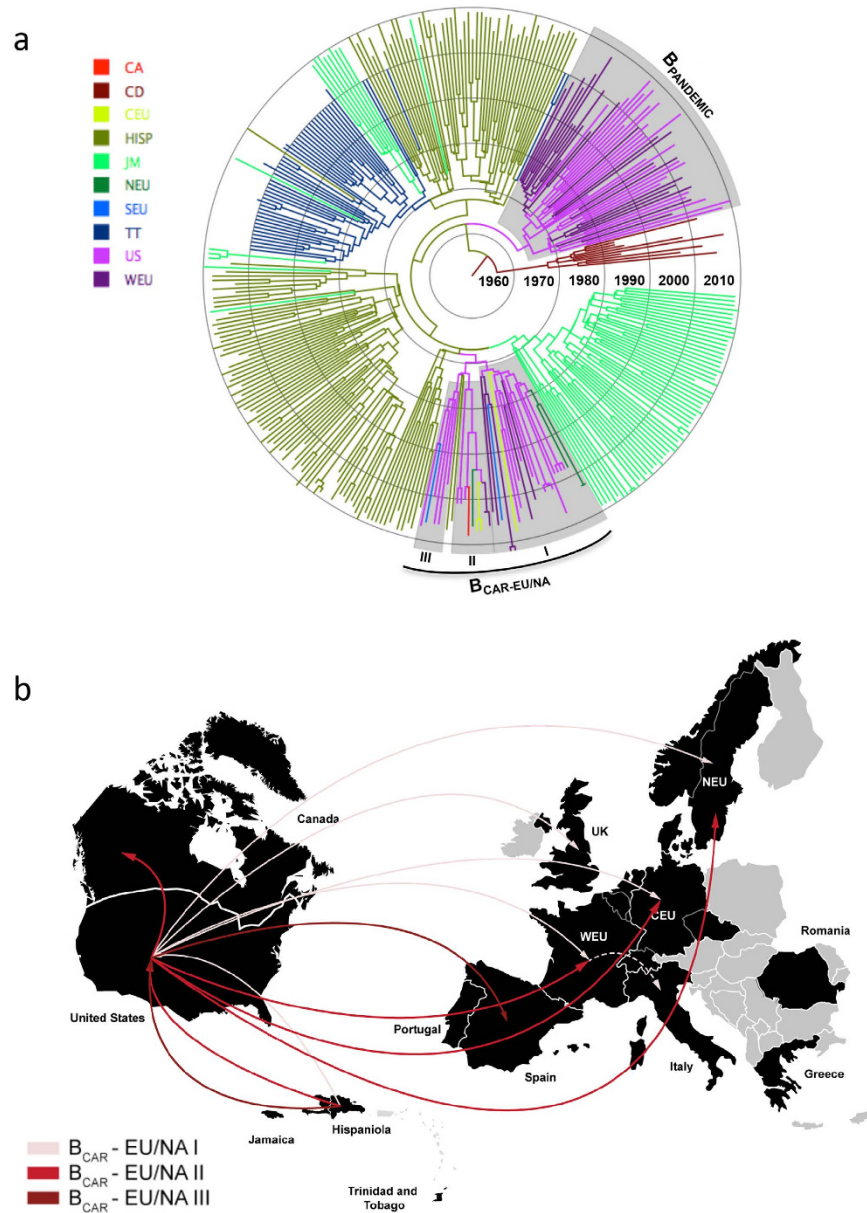


Figure 5. Spatiotemporal dynamics of dissemination of HIV-1 B_{CAR} clades from North America to Europe. (a) Time-scaled Bayesian MCMC tree of HIV-1 B_{CAR} *pol* sequences from intercontinental B_{CAR} clades ($B_{CAR-NA/EU-I}$, $B_{CAR-NA/EU-II}$ and $B_{CAR-NA/EU-III}$) combined with B_{CAR} sequences from the Caribbean ($n = 258$), $B_{PANDEMIC}$ sequences from the US and France ($n = 50$) and subtype D sequences ($n = 10$) from the DRC. Branches are colored according to the most probable location state (see Table S3) of their descendant nodes as indicated in the legend (upper left). Colored boxes indicate the positions of intercontinental B_{CAR} clades and the $B_{PANDEMIC}$ clade. Branch lengths are depicted in units of time (years). The tree was automatically rooted under the assumption of a relaxed molecular clock. See the legend of Fig. 3 for location code. (b) Lines between locations represent branches in the Bayesian MCC tree along which viral migration events occurred. Dispersion pathways of different intercontinental B_{CAR} clades are represented with different colors (see the legend at bottom left). Epidemiological links supported by Bayes factor rates >3 and <3 are displayed with continuous and discontinuous lines, respectively. See the legend of Fig. 4 for location code. Maps were created with Adobe Illustrator CC from templates obtained from d-maps.com (Caribbean: http://d-maps.com/pays.php?num_pay=118&lang=es; North America: http://d-maps.com/pays.php?num_pay=119&lang=es; and Europe: http://d-maps.com/pays.php?num_pay=192&lang=es).

Although the time-scale here reconstructed was largely determined by the informative prior distribution specified on substitution rate, the T_{MRCA} estimated for subtype B/D, subtype B and the $B_{PANDEMIC}$ clades were fully consistent with those obtained in previous studies^{12,49,50}, thus indicating that our informative prior produced reliable time-scale estimates. According to our estimations, the T_{MRCA} of most (77%) international B_{CAR} clades detected in

North America and Europe was traced to between the early 1970s and the early 1980s, whereas the onset date of a significant fraction (40%) of country-specific B_{CAR} clades detected in those regions was traced to between the late 1970s and the late 1980s. This demonstrates that despite their low prevalence, the B_{CAR} strains were introduced and circulate in North America and Europe since the early stages of the AIDS epidemic. In sharp contrast to the B_{PANDEMIC} clade, however, all B_{CAR} introductions in North America and Europe seem to have resulted in dead-end infections or in outbreaks of small size, as previously demonstrated for most Latin American countries²⁰.

One interesting question is why B_{CAR} strains introduced into North America and Europe fail to establish large secondary outbreaks. One hypothesis is that the B_{PANDEMIC} clade displays a higher transmissibility than B_{CAR} clades, but some evidences argue against it. First, analysis of partial genome regions revealed a paucity of amino acid substitutions mapping onto the branch leading to the B_{PANDEMIC} ancestor¹², suggesting that this clade probably possessed no selective advantage over B_{CAR} strains. Second, in several Caribbean countries where both pandemic and non-pandemic lineages co-circulate, the B_{CAR} clades reached a much higher prevalence than the B_{PANDEMIC} one^{19,20}. Another hypothesis is that the epidemic outcome of different subtype B strains was mostly shaped by ecological factors, like the chance of colonization of different populations. It is possible that the B_{PANDEMIC} ancestor was the only subtype B Caribbean strain to gain access to the globally interconnected populations of MSM and IDU from the US, Europe and Latin America at the early stages of the epidemic¹⁵, thus resulting in a much more efficient dissemination at both local and global level.

In conclusion, this study demonstrates that several non-pandemic HIV-1 B_{CAR} strains have been introduced from the Caribbean into North America and Europe since the early 1970s onwards. Some B_{CAR} strains were spread locally in the US, Canada and several European countries. We further detected short-distance spreading of B_{CAR} lineages between neighboring European countries and long-distance disseminations between the US and Europe. Despite their early and frequent introduction, the B_{CAR} strains only comprise a very low fraction of all HIV-1 subtype B infections from North America (3%) and Europe (1%). The epidemiological characterization of the transmission networks that sustain the dissemination of the B_{CAR} clades in North America and Europe will be of paramount importance to determine why no large B_{CAR} outbreaks have been established in those regions.

References

- UNAIDS. Report on the global AIDS epidemic, 2013. Available at http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf. (Accessed 10 August 2016).
- Centers for Disease Control and Prevention (CDC). HIV Surveillance Report, 2014. Available at: <http://www.cdc.gov/hiv/library/reports/surveillance/>. (Accessed 10 August 2016).
- Public Health Agency of Canada. Summary: Estimates of HIV incidence, prevalence and proportion undiagnosed in Canada, 2014. Available at: <http://healthycanadians.gc.ca/publications/diseases-conditions-maladies-affections/hiv-aids-estimates-2014-vih-sida-estimations/alt/hiv-aids-estimates-2014-vih-sida-estimations-eng.pdf>. (Accessed 10 August 2016).
- Pharris, A. *et al.* Trends in HIV surveillance data in the EU/EEA, 2005 to 2014: new HIV diagnoses still increasing in men who have sex with men. *Euro surveillance: European communicable disease bulletin*. **20**, (2015).
- Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS*. **25**, 679–689 (2011).
- Pyne, M. T., Hackett, J. Jr., Holzmayer, V. & Hillyard, D. R. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. *J Clin Microbiol*. **51**, 2662–2669 (2013).
- Siemieniuk, R. A., Beckthold, B. & Gill, M. J. Increasing HIV subtype diversity and its clinical implications in a sentinel North American population. *The Canadian journal of infectious diseases & medical microbiology*. **24**, 69–73 (2013).
- Stanojevic, M. *et al.* HIV1 molecular epidemiology in the Balkans: a melting pot for high genetic diversity. *AIDS Rev*. **14**, 28–36 (2012).
- Abecasis, A. B. *et al.* HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology*. **10**, 7 (2013).
- Beloukas, A. *et al.* Molecular epidemiology of HIV-1 infection in Europe: An overview. *Infect Genet Evol*. In press (2016).
- Thomson, M. M. & Najera, R. Increasing HIV-1 genetic diversity in Europe. *J Infect Dis*. **196**, 1120–1124 (2007).
- Gilbert, M. T. *et al.* The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA*. **104**, 18566–18570 (2007).
- Lukashov, V. V., Kuiken, C. L., Vlahov, D., Coutinho, R. A. & Goudsmit, J. Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. *AIDS Res Hum Retroviruses*. **12**, 1179–1183 (1996).
- Kuiken, C., Thakallapalli, R., Esklid, A. & de Ronde, A. Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am J Epidemiol*. **152**, 814–822 (2000).
- Junqueira, D. M. & Almeida, S. E. HIV-1 subtype B: Traces of a pandemic. *Virology*. **495**, 173–184 (2016).
- Hue, S., Pillay, D., Clewley, J. P. & Pybus, O. G. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci USA*. **102**, 4425–4429 (2005).
- Paraskevis, D. *et al.* Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology*. **6**, 49 (2009).
- Frentz, D. *et al.* Limited cross-border infections in patients newly diagnosed with HIV in Europe. *Retrovirology*. **10**, 36 (2013).
- Cabello, M., Mendoza, Y. & Bello, G. Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. *PLoS One*. **9**, e106045 (2014).
- Cabello, M., Junqueira, D. M. & Bello, G. Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. *AIDS*. **29**, 483–492 (2015).
- Mendoza, Y. *et al.* Human Immunodeficiency Virus Type 1 (HIV-1) Subtype B Epidemic in Panama Is Mainly Driven by Dissemination of Country-Specific Clades. *PLoS One*. **9**(4), e95360 (2014).
- Nurse, K. Diaspora, Migration and Development in the Caribbean. Policy Paper. Canadian Foundation for the Americas (FOCAL). (2004). Available at http://focal.ca/pdf/migration_Nurse_diaspora_migration_development_Caribbean_September_2004_FPP-04-6.pdf. (Accessed 10 August 2016).
- Borland, R. *et al.* HIV/AIDS and mobile populations in the Caribbean: a baseline assessment. International Organization for Migration (IOM). (2004) Available at <https://publications.iom.int/books/hiv-aids-and-mobile-populations-caribbean-baseline-assessment>. (Accessed 10 August 2016).
- Cordova Alcaraz, R. Migratory routes and dynamics between Latin American and Caribbean (LAC) countries and between LAC and the European Union. International Organization for Migration (IOM). (2012) Available at https://publications.iom.int/system/files/pdf/migration_routes_digital.pdf. (Accessed 10 August 2016).
- de Oliveira, T. *et al.* An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics*. **21**, 3797–3800 (2005).

26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
27. Berger, S. A., Krompass, D. & Stamatakis, A. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst Biol.* **60**, 291–302 (2011).
28. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **30**, 1312–1313 (2014).
29. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* **59**, 307–321 (2010).
30. Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33**, W557–W559 (2005).
31. Posada, D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* **25**, 1253–1256 (2008).
32. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* **55**, 539–552 (2006).
33. Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* **60**, 685–699 (2011).
34. Rambaut, A. FigTree v1.4: Tree Figure Drawing Tool. Available at <http://tree.bio.ed.ac.uk/software/figtree/>. (Accessed 10 August 2016)
35. Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics.* **161**, 1307–1320 (2002).
36. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* **7**, 214 (2007).
37. Suchard, M. A. & Rambaut, A. Many-core algorithms for statistical phylogenetics. *Bioinformatics.* **25**, 1370–1376 (2009).
38. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
39. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* **22**, 1185–1192 (2005).
40. Rambaut, A., Lam, T. T., Carvalho, L. M. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2**, vew007 (2016).
41. Zehender, G. *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.* **55**, 156–160 (2010).
42. Chen, J. H. *et al.* Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. *PLoS One.* **6**, e25286 (2011).
43. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* **5**, e1000520 (2009).
44. Ferreira, M. A. R. & M.A., S. Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics.* **26**, 355–368 (2008).
45. Rambaut, A. & Drummond, A. Tracer v1.6. Available at <http://tree.bio.ed.ac.uk/software/tracer/>. (Accessed 10 August 2016).
46. Huelsenbeck, J. P., Larget, B., Miller, R. E. & Ronquist, F. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst Biol.* **51**, 673–688 (2002).
47. Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics.* **27**, 2910–2912 (2011).
48. Faria, N. R. *et al.* Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. *J Gen Virol.* **93**, 889–899 (2012).
49. Junqueira, D. M. *et al.* Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One.* **6**, e27489 (2011).
50. Faria, N. R. *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science.* **346**, 56–61 (2014).
51. Caribbean Tourism Organization (CTO). Caribbean Tourism Review 2014. Available at <http://www.onecaribbean.org/statistics/annual-reviews-prospects/>. (Accessed 10 August 2016).
52. Kramer, M. A. *et al.* Migrants travelling to their country of origin: a bridge population for HIV transmission? *Sexually transmitted infections.* **84**, 554–555 (2008).
53. Kramer, M. A. *et al.* HIV transmission patterns among The Netherlands, Suriname, and The Netherlands Antilles: a molecular epidemiological study. *AIDS Res Hum Retroviruses.* **27**, 123–130 (2011).

Acknowledgements

This work was supported by Public Health Service grant 490378/2013-7 from the CNPq (Brazil) - MEC (Uruguay).

Author Contributions

The study was conceived and designed by G.B. and H.R. M.C. performed the data acquisition and determination of HIV subtype, H.R. performed phylogenetic inference and G.B. performed the phylogeographic inference. G.B. wrote the manuscript and M.C. produced all figures. All of the authors reviewed the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Cabello, M. *et al.* Multiple introductions and onward transmission of non-pandemic HIV-1 subtype B strains in North America and Europe. *Sci. Rep.* **6**, 33971; doi: 10.1038/srep33971 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016