Evolutionary history and spatiotemporal dynamics of the HIV-1 subtype B epidemic in Guatemala

Yaxelis Mendoza[1,2☯], Claudia García-Morales[3☯], Gonzalo Bello[4☯], Daniela Garrido-Rodríguez[3], Daniela Tapia-Trejo[3], Juan Miguel Pascale[1], Amalia Carolina Girón-Callejas[5], Ricardo Mendizábal-Burastero[5], Ingrid Yessenia Escobar-Urias[6], Blanca Leticia García-González[6], Jessenia Sabrina Navas-Castillo[6], María Cristina Quintana-Galindo[6], Rodolfo Pinzón-Meza[6], Carlos Rodolfo Mejía-Villatoro[6†‡], Santiago Avila-Ríos[3‡*], Gustavo Reyes-Terán[3‡*]

1 Gorgas Memorial Institute for Health Studies, Panama City, Panama, **2** Department of Genetics and Molecular Biology, University of Panama, Panama City, Panama, **3** Centre for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico, **4** Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, **5** Universidad del Valle de Guatemala, Guatemala City, Guatemala, **6** Infectious Diseases Clinic, Roosevelt Hospital, Guatemala City, Guatemala

☯ These authors contributed equally to this work.
† Deceased.
‡ Joint senior authors.
* gustavo.reyesteran@gmail.com (GRT); santiago.avila@cieni.org.mx (SAR)

Abstract

Different explanations exist on how HIV-1 subtype B spread in Central America, but the role of Guatemala, the Central American country with the highest number of people living with the virus, in this scenario is unknown. We investigated the evolutionary history and spatio-temporal dynamics of HIV-1 subtype B in Guatemala. A total of 1,047 HIV-1 subtype B *pol* sequences, from newly diagnosed ART-naïve, HIV-infected Guatemalan subjects enrolled between 2011 and 2013 were combined with published subtype B sequences from other Central American countries (n = 2,101) and with reference sequences representative of the $B_{PANDEMIC}$ and B_{CAR} lineages from the United States (n = 465), France (n = 344) and the Caribbean (n = 238). Estimates of evolutionary, demographic, and phylogeographic parameters were obtained from sequence data using maximum likelihood and Bayesian coalescent-based methods. The majority of Guatemalan sequences (98.9%) belonged to the $B_{PANDEMIC}$ clade, and 75.2% of these sequences branched within 10 monophyletic clades: four also included sequences from other Central American countries (B_{CAM-I} to B_{CAM-IV}) and six were mostly (>99%) composed by Guatemalan sequences (B_{GU} clades). Most clades mainly comprised sequences from heterosexual individuals. Bayesian coalescent-based analyses suggested that B_{GU} clades originated during the 1990s and 2000s, whereas B_{CAM} clades originated between the late 1970s and mid 1980s. The major hub of dissemination of all B_{GU}, and of B_{CAM-II}, and B_{CAM-IV} clades was traced to the Department of Guatemala, while the root location of B_{CAM-I} and $B_{CAM-III}$ was traced to Honduras. Most Guatemalan clades experienced initial phases of exponential growth (0.23 and 3.6 year^{-1}), followed by

recent growth declines. Our observations suggest that the Guatemalan HIV-1 subtype B epidemic is driven by dissemination of multiple $B_{PANDEMIC}$ founder viral strains, some restricted to Guatemala and others widely disseminated in the Central American region, with Guatemala City identified as a major hub of viral dissemination. Our results also suggest the existence of different sub-epidemics within Guatemala for which different targeted prevention efforts might be needed.

Introduction

It is commonly accepted that the HIV-1 subtype B epidemic likely moved out of Africa via a single introduction to Haiti in the mid 1960s [1]. This event was followed by a wide spread of phylogenetically distinct Caribbean-specific lineages (B_{CAR}) throughout the Caribbean islands [1–3], with non-pandemic B_{CAR} lineages accounting for more than 40% of infections in the region [3]. A pandemic B lineage ($B_{PANDEMIC}$), encompassing most subtype B infections around the world, subsequently emerged in the late 1960s [1], spreading from the Caribbean to North America and later expanding to other parts of the world [1, 4, 5], including possible re-introductions to the Caribbean [6].

The precise routes of introduction of HIV-1 subtype B to Central America remain to be clarified. One study suggests that a single early introduction of a $B_{PANDEMIC}$ strain possibly to Honduras during the 1960s originated a Central American lineage (B_{CAM}) that accounts for approximately 60% of current subtype B cases in the region [7]. Other sub-epidemics originating mostly during the 1970s and early 1980s were generated by the spread of additional founder viruses that tended to cluster according to the country of origin [6, 7]. A more detailed study in Panama, however, showed that the subtype B epidemic in this country is mostly driven by the expansion of multiple founder $B_{PANDEMIC}$ and B_{CAR} strains, that have remained mostly restricted to this country [8].

Guatemala is the country comprising the highest number of individuals estimated to be living with HIV (49,000) in Central America, with prevalence in the general population of 0.5% [9]. The Guatemalan HIV epidemic is concentrated in most-at-risk groups, with 8.9% prevalence in men who have sex with men (MSM), 23.8% in transgender women and 1.0–2.6% in female sex workers (FSW) [10]. It is estimated that men who have sex with men comprise approximately 30% of the total estimated number of persons living with HIV in Guatemala and that heterosexual transmission accounts for 68% of cases [11]. The first AIDS case in Guatemala was recorded in 1984 in a young homosexual man who had resided in the United States, and cases in women were not reported until 1986 [12]. Geographic distribution of the Guatemalan HIV epidemic reflects economic development with the involvement of departments with higher commercial activity and with marked migratory routes, including Guatemala, Escuintla, San Marcos, Retalhuleu, Quetzaltenango, Izabal, Petén and Suchitepéquez, which concentrate 76% of notified cases [12]. A higher proportion of cases is found in Ladino (Mestizo) populations (77%) compared to Maya (indigenous) populations (20%), and 60% of infected individuals are male (approximately 60% of the general population is considered Ladino and 40% indigenous) [10]. Nevertheless, the molecular epidemiological history of HIV has not been assessed in Guatemala. For rapidly evolving pathogens such as HIV, phylogenetic methods can contribute to understand the complex dynamics of viral spread at the population level. Moreover, molecular epidemiology studies have the potential to better inform targeted public health interventions to contain the HIV epidemic at the local, national, or regional

context, by identifying key clinical, demographic and geographical correlates of transmission. This information can help to focus limited resources to more effective HIV prevention efforts, including for example, the active search for new infection cases to link them to care and the roll out of pre-exposure prophylaxis [13].

The aim of the present study was to describe the spatiotemporal dynamics of dissemination of HIV subtype B in Guatemala, assessing its possible evolutionary relations with other viruses circulating in the Americas. The study was performed on a large HIV *pol* sequence dataset from individuals enrolled from 2010 to 2013 at the Roosevelt Hospital, one of the national HIV reference centers.

Materials and methods

Ethics statement

This study was evaluated and approved by the Ethics Committees of the National Institute of Respiratory Diseases (INER) in Mexico City (E06-09), and the Roosevelt Hospital in Guatemala City, and was conducted according to the principles of the Declaration of Helsinki. All participants gave written informed consent before blood sample donation.

Guatemalan HIV-1 *pol* sequences

Blood samples from 1,083 HIV-1-infected Guatemalan individuals were collected at the Infectious Diseases Clinic of the Roosevelt Hospital in Guatemala City from October 2010 to December 2013 as part of an HIV pre-treatment drug resistance study [14]. Briefly, antiretroviral treatment-naïve individuals were enrolled in a cross-sectional study according to new clinical diagnoses at spontaneous demand to the clinic, also including individuals on follow-up visits, previous to starting antiretroviral treatment. Every eligible person attending the clinic was invited to participate in the study. No exclusion criteria were applied except for previous exposure to antiretroviral drugs. After giving written informed consent, participants donated a single blood sample. The complete protease (PR) and the first part of the reverse transcriptase (RT) of the pol gene (nucleotides 2253 to 3275 of reference strain HXB2) were amplified and sequenced using an in-house developed methodology as previously described [14]. Sequences used in the study are available at S1 Data. Serving 30% of all HIV-infected persons in Guatemala, the Roosevelt Hospital is the largest of a total of 18 clinics providing antiretroviral treatment in the country [15]. Although, the Roosevelt Hospital is located in the Department of Guatemala, the persons that initiate ART in this institution come from all over the country [16], with 45% of clients from other departments [14]. The place of residence distribution of Roosevelt Hospital's clients reflects the national distribution of HIV cases [11, 16]. Also, socio-demographic characteristics (including female-to-male ratio, mean age, race, marital status and literacy) of Roosevelt Hospital's clients is similar to those of the national HIV-infected population [16, 17].

HIV-1 subtype B *pol* reference sequences

The HIV-1 subtype B *pol* sequences from Guatemala were aligned with: 1) subtype B *pol* sequences representative of the $B_{PANDEMIC}/B_{CAR}$ clades (US/France = 809/Caribbean = 238) [3, 8]; 2) Panamanian subtype B *pol* sequences representative of the country-specific clades (*n* = 583) [8]; and 3) all available subtype B *pol* sequences from other Central American countries (*n* = 694) and Mexico (*n* = 824) that matched the above selected genomic region (S1 Table). All HIV-1 subtype B *pol* reference sequences used in this study were available at Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by December 2014.

Sequence alignment and subtype assignment

Sequences were aligned using the ClustalW program implemented in Mega 6.0 software [18, 19]. The subtype assignment of all sequences included was confirmed using REGA HIV sub-typing tool v.2 [20] and by performing Neighbor-Joining (NJ) phylogenetic analyses with HIV-1 group M reference sequences downloaded from Los Alamos HIV Sequence Database. The NJ phylogenies were constructed under the Tamura-Nei evolutionary model and the reliability of tree topologies was assessed by bootstrap analysis with 500 replicates using the MEGA 6.0 software package [19]. Bootstrap values above 75% were considered significant. All sites associated 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 dataset. All alignments are available from the authors upon request.

Phylogenetic analysis

Maximum Likelihood (ML) phylogenetic trees were inferred under GTR+I+Γ_4 nucleotide substitution model selected using the jModeltest program [21]. The ML tree was reconstructed with the PhyML program [22] using an online web server [23]. 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*) [24] based on the Shimodaira-Hasegawa-like procedure. *aLRT* values above 0.8 were considered significant. The ML trees were visualized using the FigTree v1.4.0 program [25].

Analysis of spatiotemporal dispersion patterns

The evolutionary rate (nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{mrca}, years) and the spatial diffusion of HIV-1 Guatemalan clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 [26, 27] with BEAGLE [28]. Analyses were performed using the GTR+I+Γ_4 nucleotide substitution model and the temporal scale of evolutionary process was directly estimated from the sampling dates of the sequences using a relaxed uncorrelated lognormal molecular clock model [29]. Migration events throughout the phylogenetic histories were reconstructed using a reversible discrete phylogeography model [30]. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the effective sample size (ESS) and the 95% Highest Posterior Density (HPD) values respectively using the TRACER v1.5 program [31]. Maximum clade credibility (MCC) trees were summarized with TreeAnnotator v1.8.1 and visualized with FigTree v1.4.0. Migratory events were summarized using the SPREAD application [32].

Reconstruction of demographic history

The mode and rate (*r*, years^{-1}) of population growth of major HIV-1 clades circulating in Guatemala was also estimated using BEAST v1.8 software. Analyses were also performed using the GTR+I+Γ_4 nucleotide substitution model and the temporal scale of evolutionary process were estimated using a relaxed uncorrelated lognormal molecular clock model. Changes in effective population size through time were initially estimated using a Bayesian Skyline coalescent tree prior [33] and estimates of the population growth rate were subsequently obtained using the parametric model (logistic, exponential or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and

stepping-stone sampling (SS) methods [34]. MCMC chains were run for $50–500 \times 10^6$ generations. Convergence and uncertainty of parameter estimates were assessed as explained above.

Statistical analysis

Demographic and clinical variables of individuals belonging to the different clades were compared using linear regression or Chi-square test according to variable type. False Discovery Rate (FDR) for multiple comparisons was estimated with Storey q-values. The threshold for significance was set at $p < 0.05$ and $q < 0.2$. Analysis was performed using STATA/SE version 14 (StataCorp, College Station, TX).

Results

Most HIV-1 subtype B Guatemalan sequences belong to the $B_{PANDEMIC}$ clade

Of the 1,083 HIV-1 *pol* Guatemalan sequences here analyzed, we selected 1,047 subtype B sequences geographically distributed among the 22 departments of Guatemala. The remaining 36 sequences were excluded because of possible quality issues (unusual indel mutations/rare polymorphisms) ($n = 21$) or non-B subtype classification ($n = 15$). In order to better understand the origin of the Guatemalan HIV-1 subtype B epidemic, Guatemalan sequences were combined with representative sequences of the $B_{PANDEMIC}$ and B_{CAR} lineages (S1 Table) to perform a phylogenetic analysis. The ML analysis revealed that, as expected, the B_{CAR} sequences occupied the deepest branches within the subtype B phylogeny; whereas the $B_{PANDEMIC}$ sequences branched as a well-supported ($aLRT = 0.88$) monophyletic subgroup nested within the B_{CAR} clades (Fig 1). Of the 1,047 sequences analyzed from Guatemala, 1,036 (98.9%) branched within the $B_{PANDEMIC}$ clade and only 11 sequences (1.1%) branched among the B_{CAR} lineages (Fig 1). This analysis also revealed that 787 (75.2%) of the Guatemalan sequences branched in 10 country-specific monophyletic clades of large size ($29 \leq n \leq 209$; B_{GU-I} to B_{GU-X}), 226 (21.6%) branched in 43 country-specific clusters of small size ($2 \leq n \leq 17$), and the remaining 34 (3.2%) represented non-clustered sequences (Fig 1).

Identification of country- and region-specific HIV-1 subtype B clades

HIV-1 subtype B sequences from the largest Guatemalan clades here identified were next aligned with subtype B sequences from Mexico, Panama and other Central American countries (S1 Table). ML analysis revealed that all sequences from Panama and Costa Rica and most sequences from Mexico (98%) branched outside the major Guatemalan clades (Figs 2 and 3). By contrast, a large proportion of sequences from Honduras (92%), El Salvador (54%), and Belize (22%) branched within the previously defined Guatemalan clades B_{GU-IV}, B_{GU-VII}, B_{GU-IX} and B_{GU-X} that were thus reclassified as Central American clades B_{CAM-IV}, B_{CAM-II}, $B_{CAM-III}$ and B_{CAM-I}, respectively (Fig 2). The prevalence of each major Central American clade greatly varied among countries (S2 Table). In Guatemala, the predominant clade was B_{CAM-I} (19%), followed by clades B_{CAM-II} (12%), B_{CAM-IV} (10%) and $B_{CAM-III}$ (8%). In Honduras, the most prevalent clade was B_{CAM-I} (73%), followed by clades $B_{CAM-III}$ (12%) and B_{CAM-II} (7%). In El Salvador, the predominant clade was B_{CAM-II} (34%), followed by clades B_{CAM-IV} (12%) and B_{CAM-I} (8%). In Belize, B_{CAM-I} and B_{CAM-II} clades seemed to circulate with similar prevalence (11%). Guatemalan clades (B_{GU-I}, B_{GU-II}, B_{GU-III}, B_{GU-V}, B_{GU-VI} and $B_{GU-VIII}$) were mostly (>99%) composed by sequences from that country and were thus classified as country-specific clades. Among subtype B sequences from Guatemala included in the study ($n = 1,047$), 49.3% ($n = 516$) belonged to Central America-specific clades and 23.8% ($n = 249$) belonged to Guatemala-specific clades.
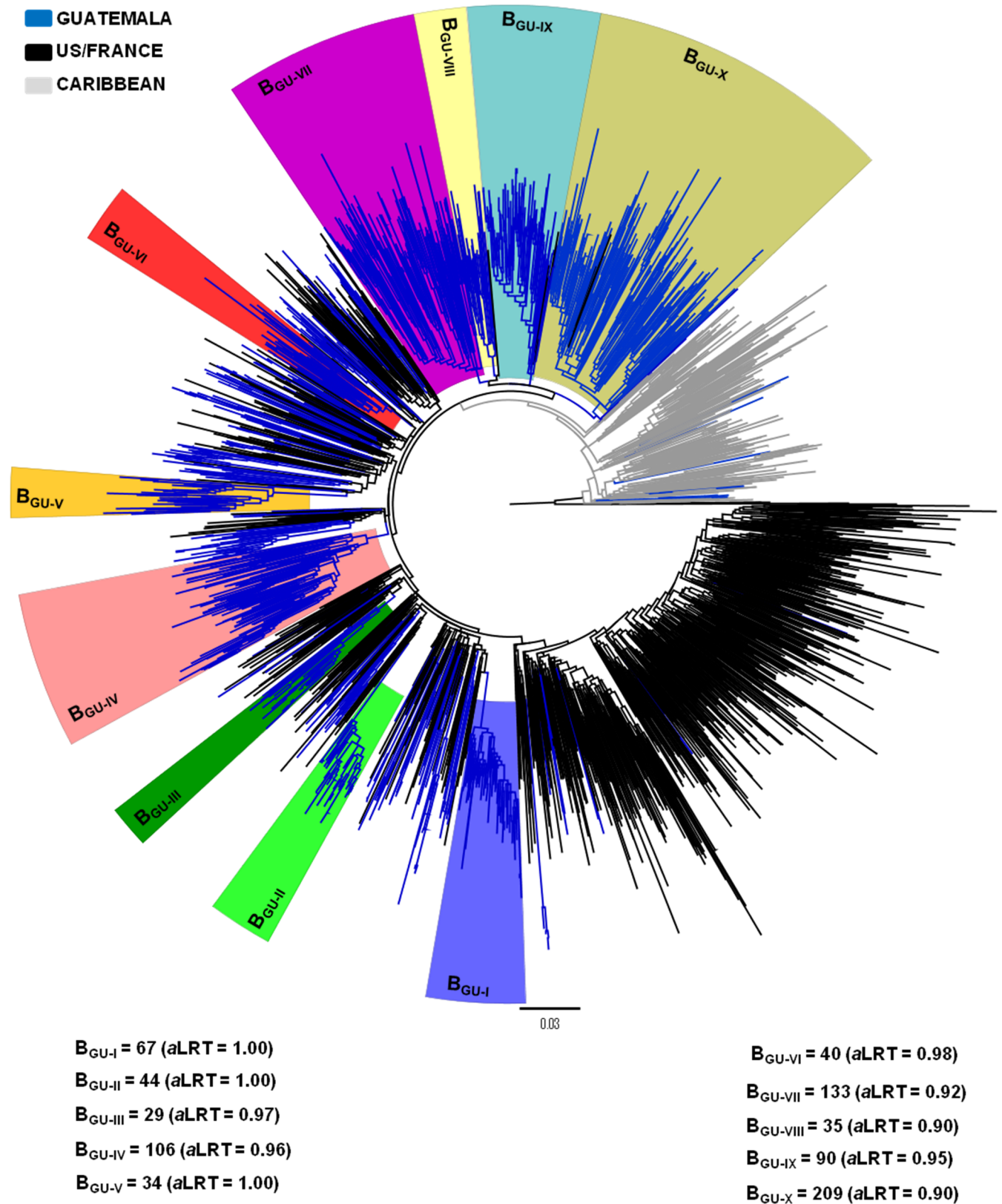
Fig 1. Maximum likelihood phylogenetic tree of HIV-1 subtype B *pol* (~1000 pb) sequences circulating in Guatemala (*n* = 1047), and representative sequences of the B$_{PANDEMIC}$ (US = 465, France = 344) and the B$_{CAR}$ (Caribbean = 238) clades. Branches are colored according to the geographic origin of

each sequence, as indicated in the legend (top left). Colored boxes highlight the position of the ten major Guatemalan HIV-1 subtype B clades (B_{GU-I} to B_{GU-X}). The number of sequences and the *a*LRT support values for each clade are indicated at the bottom. The tree was rooted using HIV-1 subtype D reference sequences. Branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

https://doi.org/10.1371/journal.pone.0203916.g001

Epidemiological characteristics of Guatemalan subjects across major HIV-1 subtype B clades circulating in Central America

We then analyzed the demographic and clinical characteristics of individuals included in the different Guatemalan and Central American clades (Table 1). Significant differences in all the variables analyzed were observed between the B_{GU-II} clade and the other B_{CAM} and B_{GU} lineages. The B_{GU-II} clade was mostly transmitted among young (median age 27 years), recently infected (59%), MSM (46%), with higher education level (71% with bachelor degree and 16% with high-school degree) and employment (64%) rates. In sharp contrast, all other B_{CAM} and B_{GU} clades were mostly composed of older (median age >31 years), heterosexual (>79%) individuals, with long-standing infection (>78%), and lower education (degree <21%) and employment (<53%) rates than those observed in individuals infected by the B_{GU-II} clade. Guatemala City was the residence for most (>48%) individuals of major B_{CAM} and B_{GU} clades, with exception of B_{GU-V} (24%) and $B_{CAM-III}$ (34%). The B_{GU-V} clade stands out due to a lower mean CD4 T cell count. This clade was composed by persons from outside the Department of Guatemala (77%), with a significantly lower level of education (35% with none and 53% with elementary degree) and heterosexual risk of transmission. Late arrival to clinical care in populations with lower education and socio-economic status is a common feature of many Latin American countries [35]. This may explain the common feature of low CD4 T cell counts at the time of HIV diagnosis.

Spatiotemporal dissemination of major HIV-1 subtype B clades circulating in Central America

The median estimated evolutionary rates at HIV-1 *pol* gene for major Guatemalan and Central American HIV-1 clades ranged from 2.1 x 10^{-3} subst./site/year to 2.9 x 10^{-3} subst./site/year with a great overlap of 95% HPD intervals, whereas the coefficient of variation ranged from 0.33 (0.14–0.50) to 0.55 (0.34–0.76), thus supporting the use of a relaxed molecular clock model (Table 2). The HIV-1 subtype B Guatemalan clades B_{GU-III}, B_{GU-V}, B_{GU-VI} and $B_{GU-VIII}$ arose between the early and the middle 1990s, followed by clades B_{GU-I} and B_{GU-II} between the early and the middle 2000s (Table 2). The phylogeographic analysis showed that the department of Guatemala was the most probable epicenter of dissemination of all those Guatemalan clades (*PSP*≥0.86). B_{GU-I} was the most widely disseminated lineage, being detected in 15 out of the 22 Guatemalan departments, followed by clade B_{GU-V} detected in 11 departments and El Salvador; clade $B_{GU-VIII}$ detected in 11 departments, clade B_{GU-VI} in 10 departments; clade B_{GU-II} in seven departments, and clade B_{GU-III} in six departments and Mexico (Fig 4). The departments of Alta Verapaz, Quiche, Sacatepequez and Santa Rosa were also detected as secondary disseminating locations of clades B_{GU-V}, B_{GU-I}/B_{GU-II}/$B_{GU-VIII}$, B_{GU-III} and B_{GU-II}, respectively (Fig 4).

The root location of Central American clades B_{CAM-I} and $B_{CAM-III}$ was most probably traced to Honduras (Posterior State Probability, *PSP* ≥0.99) around the middle 1980s (Table 2). The most widely disseminated viral clade, $B_{CAM-I,}$ migrated from Honduras to the department of Guatemala, El Salvador and Mexico during the early 1990s and later moved to other Guatemalan departments around 2010 (Fig 5). From the department of Guatemala, clade B_{CAM-I} was disseminated to other 16 Guatemalan departments and also to Mexico,
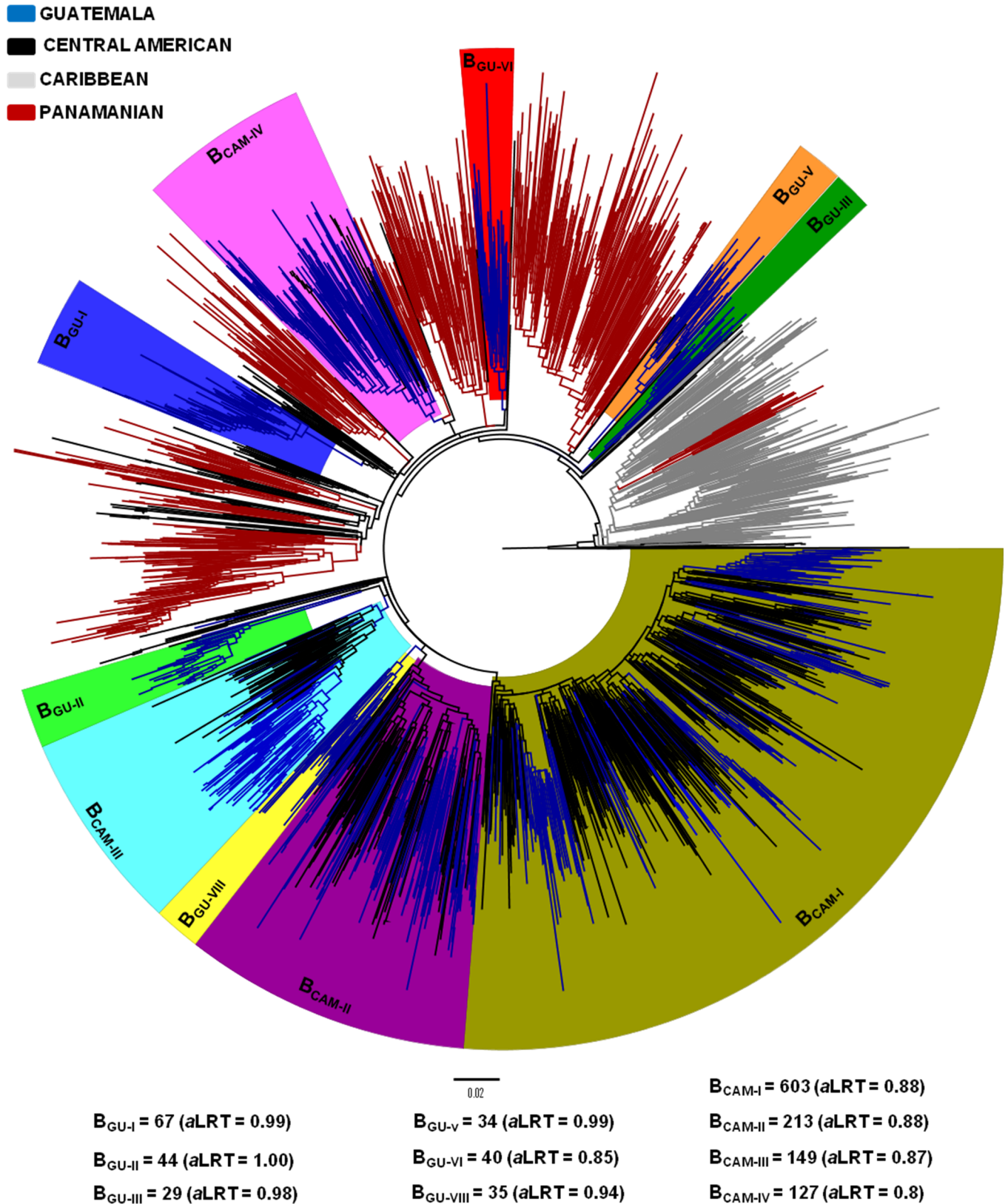
Fig 2. Maximum likelihood phylogenetic tree of HIV-1 subtype B *pol* **(~1000 pb) sequences belonging to the major Guatemalan clades (B_{GU-I} to B_{GU-X}, n = 787), and reference sequences from various Central American countries (*n* = 694), Panama (*n* = 583), and the Caribbean (*n* = 238).** Branches are

B_{GU-I} = 67 (aLRT = 0.99) B_{GU-V} = 34 (aLRT = 0.99) B_{CAM-I} = 603 (aLRT = 0.88)

B_{GU-II} = 44 (aLRT = 1.00) B_{GU-VI} = 40 (aLRT = 0.85) B_{CAM-II} = 213 (aLRT = 0.88)

B_{GU-III} = 29 (aLRT = 0.98) $B_{GU-VIII}$ = 35 (aLRT = 0.94) $B_{CAM-III}$ = 149 (aLRT = 0.87)

B_{CAM-IV} = 127 (aLRT = 0.8)

colored according to the geographic origin of each sequence, as indicated in the legend (top left). Colored boxes highlight the position of the six major Guatemalan HIV-1 subtype B clades (B_{GU}s) and the four Central American clades (B_{CAM-I} to B_{CAM-VI}). The number of sequences and *a*LRT support values for each clade are indicated at the bottom. The tree was rooted using HIV-1 subtype D reference sequences. Branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

Belize, and back to Honduras. El Salvador and the Guatemalan department of Escuintla were also secondary hubs of dissemination of clade B_{CAM-I}. Clade $B_{CAM-III}$ moved from Honduras to several Guatemalan departments (Guatemala, Zacapa, Sacatepéquez, Izabal and Chiquimula) and to Mexico since 1997 (Fig 5). The department of Guatemala acted as a secondary hub of dissemination of $B_{CAM-III}$ sending viruses to other Guatemalan departments from 2001 onwards.

The root location for clades B_{CAM-II} and B_{CAM-IV} was most probably traced to the department of Guatemala (*PSP* = 0.47 and 0.99, respectively) between the late 1970s and the early 1990s (Table 2). Clade B_{CAM-II} moved from the department of Guatemala to Honduras and El Salvador around the late 1970s, to the Guatemalan departments of Solola and Chiquimula during the 1990s, and to other 11 Guatemalan departments, Belize and Mexico since the early 2000 (Fig 5). Secondary hubs of dissemination of clade B_{CAM-II} were El Salvador from where the virus has spread to Honduras, back to the department of Guatemala, and to other seven Guatemalan departments since the early 1990s, and Honduras from where the virus has spread back to the department of Guatemala and Mexico since the early 1980s. The department of Guatemala was the major epicenter of clade B_{CAM-IV} moving from the capital to other 15 Guatemalan departments, as well as to El Salvador and Mexico since the middle 1980s. From El Salvador, clade B_{CAM-IV} moved back to the department of Guatemala.

Demographic history of major HIV-1 subtype B clades circulating in Guatemala

To reconstruct the population dynamics of the HIV-1 subtype B epidemic in Guatemala, we used the Guatemalan sequences from those clades that arose in Guatemala (B_{GU-I}, B_{GU-II}, B_{GU-III}, B_{GU-V}, B_{GU-VI} and $B_{GU-VIII}$, and B_{CAM-IV}) and well supported (aLTR $\geq$80) Guatemala-specific sub-clades within clades B_{CAM-I} ($B_{CAM-I/GU-I}$ = 51 sequences and $B_{CAM-I/GU-II}$ = 37) and $B_{CAM-III}$ ($B_{CAM-III/GU}$ = 71 sequences). Bayesian skyline plot (BSP) analyses suggest that most clades experienced an initial phase of exponential growth followed by a decline in growth rate during the 2000s, leading to stabilization of the effective population size (Ne) (Fig 6). Some clades (B_{GU-II}, B_{GU-III}, B_{CAM-IV} and $B_{CAM-I/GU-II}$), however, seem to have experienced more complex dynamics characterized by phases of expansion interleaved by periods of slow growth (Fig 6). Model comparison indicated that the best-fit demographic model was the logistic for clades B_{GU-II}, B_{CAM-IV}, B_{GU-VI}, $B_{CAM-I/GU-I}$ and $B_{CAM-I/GU-II}$, while both the logistic and exponential models fit equally well the demographic signal in clades B_{GU-I}, B_{GU-III}, B_{GU-V}, $B_{GU-VIII}$ and B_{GU-IX} (S3 Table). According to these parametric models, the mean estimated growth rate of major clades circulating in Guatemala ranged from 0.23 year^{-1} for clade B_{GU-III} to 3.59 year^{-1} for clade B_{GU-II} (Table 3).

Discussion

The HIV-1 subtype B epidemic in Guatemala is mostly related to the $B_{PANDEMIC}$ clade (98.9%), as described for the majority of the Central [7, 8, 36], North, and South [3, 37] American countries. The proportion of Guatemalan HIV-1 subtype B viruses belonging to the B_{CAR} clades (1.1%) was similar to the prevalence reported for Honduras, El Salvador, and Mexico
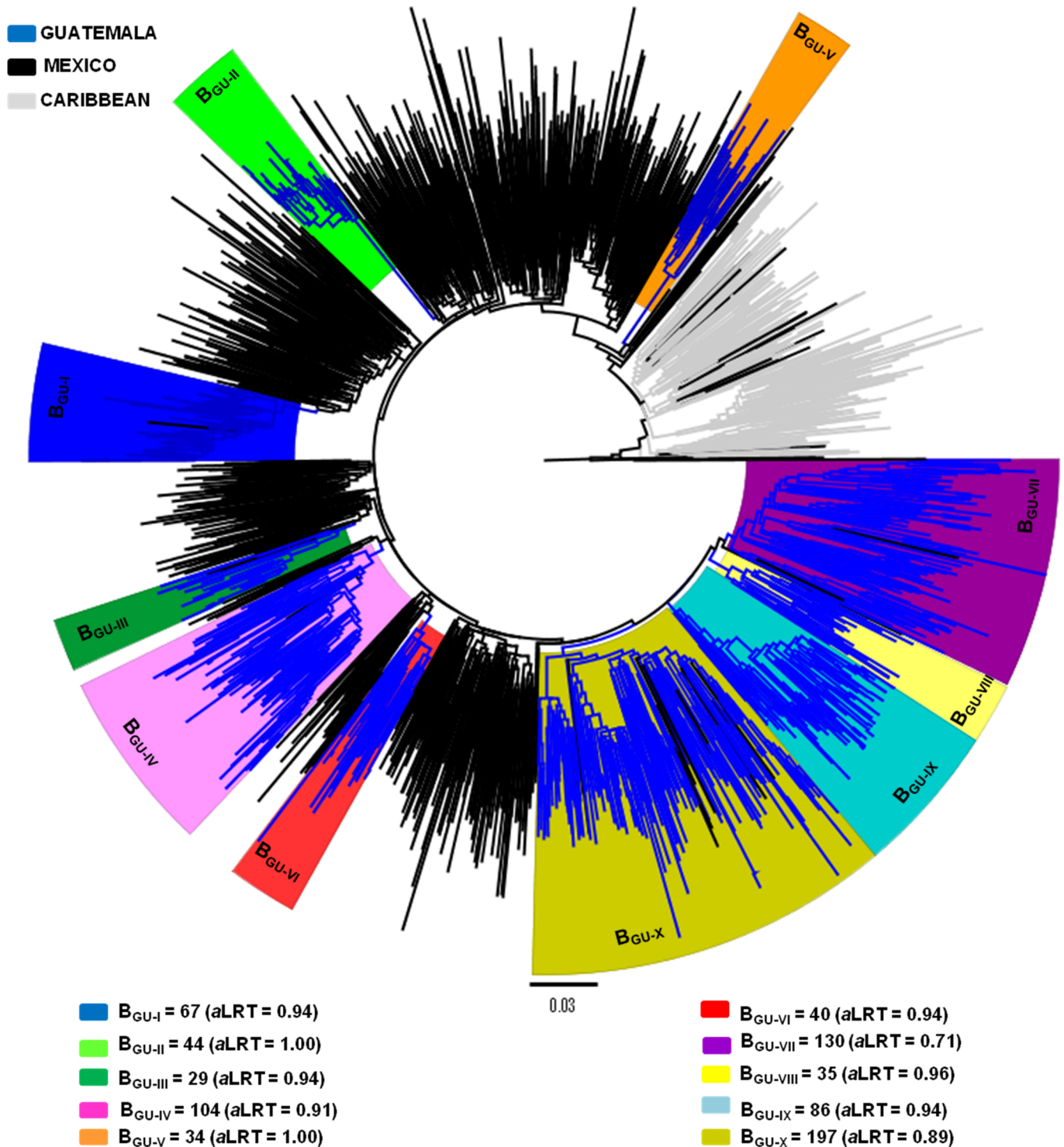
Fig 3. Maximum likelihood phylogenetic tree of HIV-1 subtype B *pol* (~1000 bp) sequences belonging to the major Guatemalan clades (B$_{GU-I}$ to B$_{GU-X}$, n = 787), and reference sequences from Mexico (*n* = 824) and the Caribbean (*n* = 238). Branches are colored according to the geographic origin of each sequence, as indicated in the legend (top left). Branches from clades B$_{GU-I}$ to B$_{GU-X}$ are colored according to clade as indicated in the legend (bottom). The tree was rooted using HIV-1 subtype D reference sequences. Branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

https://doi.org/10.1371/journal.pone.0203916.g003

Table 1. Clinical and demographic characteristics of individuals included in the large Guatemalan and Central American clusters.

	All clusters	B GU-I	B GU-II	B GU-III	B GU-V	B GU-VI	B GU-VIII	B CAM I	B CAM II	B CAM III	B CAM IV
n	765	67	44	29	34	40	35	197	129	86	104
%	100	8.8	5.8	3.8	4.4	5.2	4.6	25.8	16.9	11.2	13.6
Age (years)											
Median	34	37	**27**	**40**	31	36	36	35	35	33	34
IQR	28–44	29–46	**22–30**	**31–51**	28–42	29–43	29–48	30–44	29–45	27–44	29–43
Gender (%)											
Female	43.9	52.2	**6.8**	44.8	44.1	52.5	**65.7**	45.2	45	46.5	37.5
Male	56.1	47.8	**93.2**	55.2	55.9	47.5	**34.3**	54.8	55	53.5	62.5
Civil Status (%)											
Single	38.8	29.9	**81.8**	55.2	29.4	32.5	42.9	35	**31**	**38.4**	43.3
Married	24.2	**34.3**	**9.1**	24.1	**38.2**	22.5	22.9	**18.3**	29.5	23.3	26
Domestic Partnership	27.6	26.9	**6.8**	13.8	26.5	**35**	28.6	35	31.8	25.6	20.2
Divorced	1.31	1.49	0	0	0	2.5	0	1.52	0.78	1.16	2.88
Widow	8.1	7.5	2.3	6.9	5.9	7.5	5.7	10.2	7	11.6	7.7
Education (%)											
None	19	26.9	**4.6**	24.1	**35.3**	22.5	22.9	15.2	24.8	17.4	**11.5**
Elementary	50.6	56.7	**9.1**	51.7	52.9	52.5	54.3	54.3	52.7	53.5	49
High-school	15.6	9	15.9	3.4	8.8	12.5	20	16.2	13.2	**24.4**	19.2
Degree	14.9	7.5	**70.5**	20.7	**2.9**	12.5	**2.9**	14.2	**9.3**	**4.7**	20.2
Employment (%)											
Employed	46	49.3	**63.6**	51.7	52.9	37.5	34.3	42.6	50.4	43	43.3
Unemployed	54	50.7	**36.4**	48.3	47.1	62.5	65.7	57.4	49.6	57	56.7
Recency of infection (%)											
Recent	18.3	13.4	**59.1**	6.9	8.8	7.5	20	16.2	15.5	22.1	18.3
Long-standing	81.7	86.6	**40.9**	93.1	91.2	92.5	80	83.8	84.5	77.9	81.7
HIV risk factor (%)											
Heterosexual	87.6	**97**	**38.6**	86.2	94.1	90	94.3	**93.4**	90.7	91.9	**78.9**
MSM	8.2	**0**	**45.5**	6.9	2.9	5	2.9	**3.6**	7.8	**2.3**	**17.3**
Bisexual	3.3	1.5	15.9	6.9	0	5	2.9	2	1.6	3.5	2.9
PWID	0.5	1.5	0	0	0	0	0	0.5	0	1.2	1
Other/Unknown	0.4	0	0	0	**2.9**	0	0	0.5	0	1.2	0
Residence (%)											
Guatemala City	53.1	49.3	**77.3**	**79.3**	**23.5**	52.5	48.6	55.3	48.1	**33.7**	**67.3**
Not Guatemala City	46.9	50.7	**22.7**	**20.7**	**76.5**	47.5	51.4	44.7	51.9	**66.3**	**32.7**
Viral Load (log copies/mL)											
Mean	4.86	**5.1**	4.89	4.96	4.82	4.87	4.86	4.85	4.74	**5.09**	**4.67**
IQR	4.34–5.37	**4.47–5.62**	4.44–5.26	4.59–5.47	4.47–5.18	4.39–5.33	4.16–5.38	4.23–5.38	4.24–5.15	**4.44–5.53**	**4.10–5.13**
CD4+ T cell count (cells/μL)											
Mean	167	146	**399**	157	**80**	142	214	156	139	172	178
IQR	66–346	81–327	**256–481**	62–330	**29–206**	52–297	90–362	61–353	66–309	65–292	82–388

Significant values (each cluster vs. all the other clusters; p<0.5, q<0.2) are shown in bold.

(<1.1%) [37], but lower than the one reported in Panama (5.5%), the country where B_{CAR} clade circulation was first demonstrated in Central America [8].

Our phylogeographic analysis identified both country- and region-specific HIV-1 subtype B clades involved in the Guatemalan HIV epidemic, contrasting with the Panamanian

Table 2. Evolutionary rate and time-scale estimates of B_{GU} and B_{CAM} clades.

Clade	N	Sampling interval	μ (subst./site/year)[a]	Variation Coefficient[a]	T_{MRCA}[a]
B_{GU-I}	67	2010–2013	2.2×10^{-3} (1.5×10^{-3}–2.9×10^{-3})	0.55 (0.34–0.76)	2000 (1994–2004)
B_{GU-II}	44	2010–2013	2.4×10^{-3} (1.7×10^{-3}–3.0×10^{-3})	0.37 (0.004–0.65)	2005 (2002–2007)
B_{GU-III}	29	2010–2013	2.1×10^{-3} (1.5×10^{-3}–2.8×10^{-3})	0.42 (0.07–0.75)	1991 (1980–2000)
B_{GU-V}	34	2010–2013	2.0×10^{-3} (1.5×10^{-3}–2.8×10^{-3})	0.43 (0.23–0.65)	1992 (1983–2000)
B_{GU-VI}	40	2010–2013	2.2×10^{-3} (1.5×10^{-3}–2.9×10^{-3})	0.33 (0.14–0.50)	1994 (1986–2001)
$B_{GU-VIII}$	35	2010–2013	2.2×10^{-3} (1.5×10^{-3}–2.9×10^{-3})	0.46 (0.24–0.68)	1994 (1986–2001)
B_{CAM-I}	509	2001–2013	2.9×10^{-3} (2.8×10^{-3}–3.0×10^{-3})	0.33 (0.29–0.37)	1987 (1984–1989)
B_{CAM-II}	224	2003–2013	2.3×10^{-3} (1.7×10^{-3}–2.9×10^{-3})	0.21 (0.14–0.28)	1978 (1968–1986)
$B_{CAM-III}$	150	2001–2013	2.9×10^{-3} (2.7×10^{-3}–3.0×10^{-3})	0.31 (0.22–0.41)	1987 (1984–1990)
B_{CAM-IV}	126	2005–2013	1.9×10^{-3} (1.5×10^{-3}–2.6×10^{-3})	0.30 (0.22–0.39)	1980 (1971–1990)

[a]95% Higher Posterior Density intervals are shown in parentheses.

https://doi.org/10.1371/journal.pone.0203916.t002

epidemic characterized mostly by the dissemination of country-specific clades [8]http://www.ncbi.nlm.nih.gov/pubmed/?term=Uribe-Salas%20F%5BAuthor%5D&cauthor=true&cauthor_uid=12616149. Our study further reveals the existence of at least four major regional-specific clades in Central America (BCAM-I—B_{CAM-IV}), contrasting with the previous observation of a single HIV subtype B linage (B_{CAM} clade) accounting for most of the HIV epidemic burden in Central America [7]. The fact that the Central American clades BCAM-I—B_{CAM-IV} differ in root location (B_{CAM-I} and $B_{CAM-III}$ traced to Honduras, and B_{CAM-II} and B_{CAM-IV} traced to Guatemala), also refute the Murillo *et al* assertion that a single viral introduction event was responsible for most of the HIV-1 subtype B infections in Central America. This change in the previously described HIV-1 epidemiological history of subtype B in Central America is expected, considering the fact that the highest number of individuals estimated to be living with HIV in Central America are from Guatemala [9], and this country was not included in previous analyses.

Nearly half of all Guatemalan HIV-1 subtype B sequences included in our study branched within Central American $B_{PANDEMIC}$ clades (67%), indicating an unrestricted viral flow between main transmission networks from Guatemala and the neighboring Central American countries of Honduras and El Salvador. This result is consistent with a previous study that suggested an HIV-1 subtype B epidemiological link between the Guatemalan neighboring countries Belize, Honduras and El Salvador [6]. A much more restricted viral flow, by contrast, was observed between Guatemala and Mexico (despite their close geographic proximity) and between Guatemala and Panama. Viral flux between Guatemala and Belize, Nicaragua and Costa Rica was not possible to estimate because of the paucity of HIV-1 *pol* sequences from these countries.

About one fourth of all Guatemalan HIV-1 subtype B sequences included in our study branched within six country-specific clades of large size (B_{GU-I}, B_{GU-II}, B_{GU-III}, B_{GU-V}, B_{GU-VI}, and $B_{GU-VIII}$). The analysis of clinical and demographic characteristics, however, reveals no major difference between B_{CAM} and B_{GU} clades, with the exception of clade B_{GU-II} that differed from all other clades mostly comprising mainly young, recently infected MSM with higher education and employment rates. In the last 10 years, HIV prevalence among MSM has remained higher than 8% in Guatemala, with most of the cases (75%) related to MSM younger than 36 years with higher education [38]. Major differences, by contrast, were detected between B_{CAM} and B_{GU} clades, with regards to the estimated origin of the different lineages: B_{GU} clades displayed a much more recent median T_{MRCA} (between the 1990s and 2000s) than
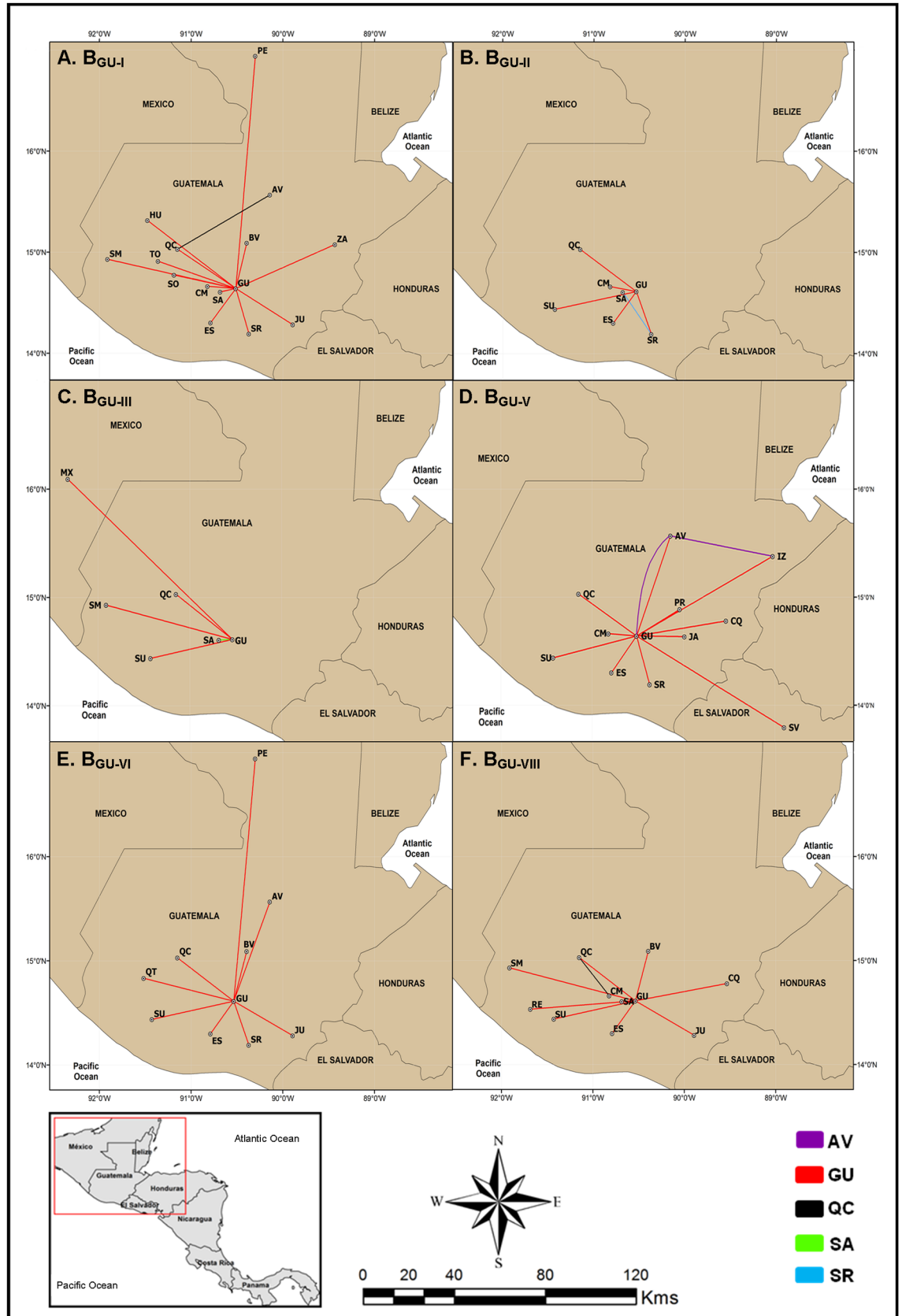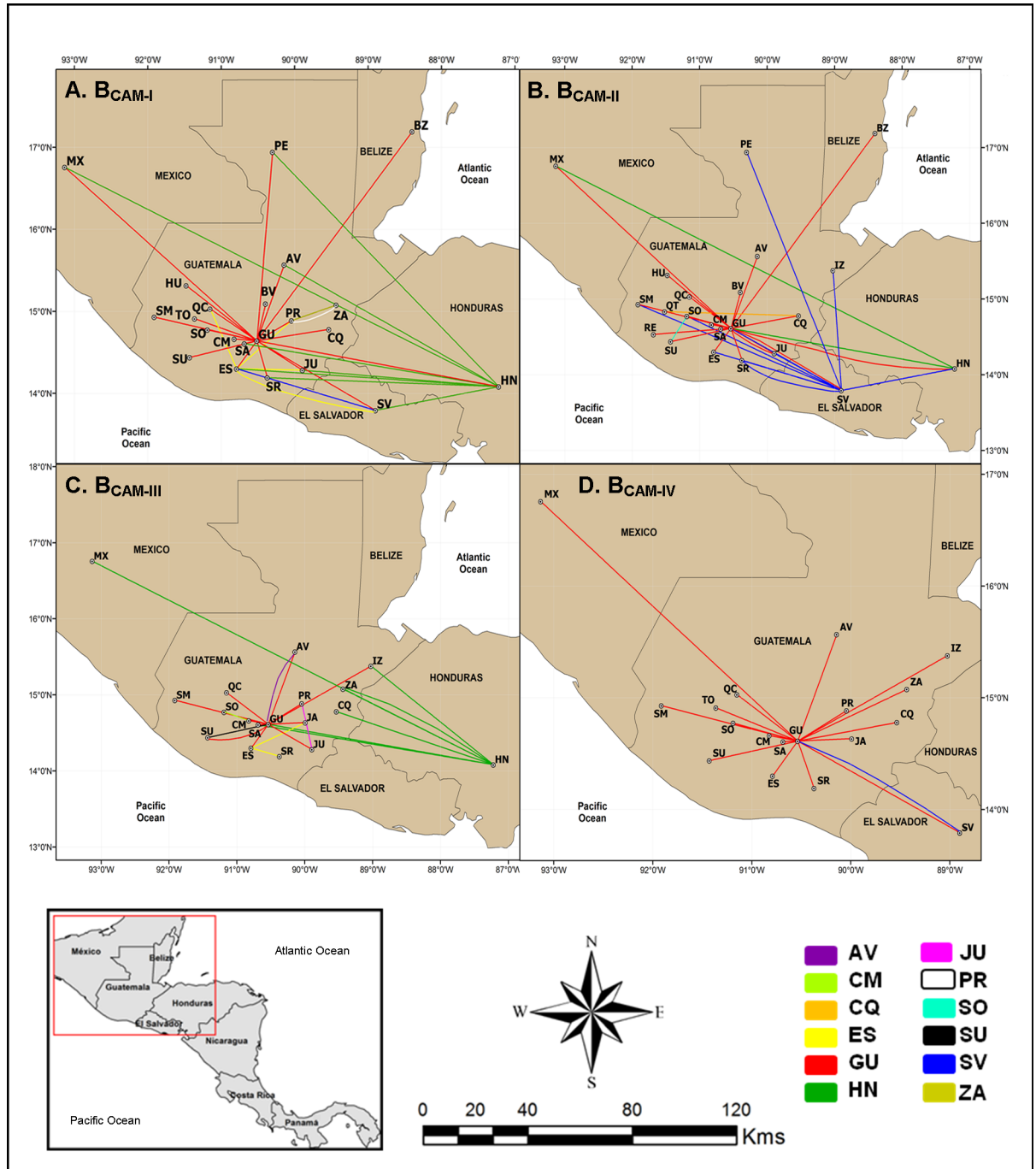
Fig 4. Spatial dissemination of HIV-1 country-specific Guatemalan clades (B_{GU}). Lines between departments and neighboring countries represent branches in the Bayesian MCC tree along which location transitions occur. Lines were colored according to the source location as indicated in the legend at the bottom. Dashed lines indicate transitions associated to nodes with low location probability support (<0.60). Guatemalan Departments: Alta Verapaz (AV), Baja Verapaz (BV), Chimaltenango (CM), Chiquimula (CQ), El Progreso (PR), Escuintla (ES), Guatemala (GU), Huehuelenango (HU), IZABAL (IZ), Jalapa (JA), Jutiapa (JU), PETEN (PE), Quetzaltenango (QT), Quiche (QC), Retalhuleu (RE), Sacatepequez (SA), San Marcos (SM), Santa Rosa (SR), Solola (SO), Suchitepequez (SU), Totonicapan (TO) and Zacapa (ZA).

the B_{CAM} clades (between the 1970s and 1980s), which may have had a great impact in the geographical expansion of each lineage.

The spatiotemporal dissemination analysis of B_{GU} and B_{CAM} clades evidenced complex migration patterns characterizing the HIV-1 epidemic in the region, involving Guatemalan departments and the neighboring countries Honduras, El Salvador, Belize and Mexico. Guatemalan internal and external migration has been attributed mostly to economic factors, with population mobility from poor rural areas to agriculture or industrial developing areas [39], and the country's civil war that lasted 32 years, from 1960 to 1996 [40]. The department of Guatemala and, to a lesser extent, Escuintla, Izabal and Petén have been the main destinations for internal-migration from the eastern and western departments of the country, with dynamic bidirectional population mobility [41, 42]. Regarding external migration, 50% of immigrants to Guatemala correspond to people of other Central American countries, mostly from El Salvador, Nicaragua and Honduras [42], yet the Guatemalan net migration rate is negative [43] with emigration flows mostly to the United States of America, and the neighboring countries of Mexico, Honduras, El Salvador and Belize [42]. Our phylogeographic reconstructions revealed that the department of Guatemala was the dissemination epicenter of the B_{GU}, B_{CAM-II} and B_{CAM-IV} clades to the rest of Guatemalan departments and to the Guatemalan neighboring countries. Additionally, the departments of Guatemala and Escuintla served as secondary hubs of dissemination of the B_{CAM-I} and $B_{CAM-III}$ clades, which emerged in Honduras, to other Guatemalan departments and to the four bordering countries in the case of B_{CAM-I}. The estimated T_{MRCA} of B_{GU} clades, ranging between the early 1990s and the middle 2000s, also coincides with an increased repatriation and deportation rate of Guatemalans from Mexico and the United States of America, respectively [42, 44], although we found no significant evolutionary links between B_{GU} clades and the Mexican subtype B epidemic.

The demographic analysis of the B_{GU-I}, B_{GU-V}, B_{GU-VI}, $B_{GU-VIII}$, and $B_{CAM-I/GU-I}$ clusters suggested that, even though all these clades had an initial phase of exponential growth, the dissemination of these clades declined during the 2000s. On the other hand, the B_{GU-II}, B_{GU-III}, B_{CAM-IV} and $B_{CAM-I/GU-II}$ clades evidenced a more complex growth dynamics with phases of expansion interleaved by periods of slow growth. The B_{GU-II} clade, characterized mostly by a population of young MSM, evidenced the highest mean estimated logistic growth rate (3.6 year^{-1}) among the major clades circulating in Guatemala. The B_{GU-II} clade growth rate was remarkably higher than those previously estimated for other country-specific B clades mainly or exclusively restricted to MSM (ranging from 0.5 to 1.6 year^{-1}) [45–48]. Furthermore, the proportion of strains that corresponds to new infections within the B_{GU-II} clade was higher (59.1%) compared with the rest of the B_{GU} and B_{CAM} clades (ranging from 6.9% of B_{GU-III} to 22.1% of $B_{CAM-III}$). As the HIV epidemic in Guatemala is concentrated in high-risk key populations, with 8.9% prevalence in MSM [10], an increase in size and proportion of the B_{GU-II} clade would be expected in the current Guatemalan HIV epidemic. These results, however, should be interpreted with caution because the exponential growth model, that was equally supported as the logistic one, showed a much lower mean growth rate (0.4 year^{-1}) for the B_{GU-II} clade. Despite the fact that other major subtype B clades in Guatemala mostly circulate

Fig 5. Spatial dissemination of HIV-1 Central American clades circulating in Guatemala (B$_{CAM-I}$ to B$_{CAM-IV}$). Lines between departments and neighboring countries represent branches in the Bayesian MCC tree along which location transitions occur. Lines were colored according to the source location as indicated in the legend at the bottom. Dashed lines indicate transitions associated to nodes with low location probability support (<0.60). Countries: Belize (BZ), El Salvador (SV), Mexico (MX) and Honduras (HN). Guatemalan Departments: Alta Verapaz (AV), Baja Verapaz (BV), Chimaltenango (CM), Chiquimula (CQ), El Progreso (PR), Escuintla (ES), Guatemala (GU), Huehuelenango (HU), IZABAL (IZ), Jalapa (JA), Jutiapa (JU), PETEN (PE), Quetzaltenango (QT), Quiche (QC), Retalhuleu (RE), Sacatepequez (SA), San Marcos (SM), Santa Rosa (SR), Solola (SO), Suchitepequez (SU), Totonicapan (TO) and Zacapa (ZA).

https://doi.org/10.1371/journal.pone.0203916.g005
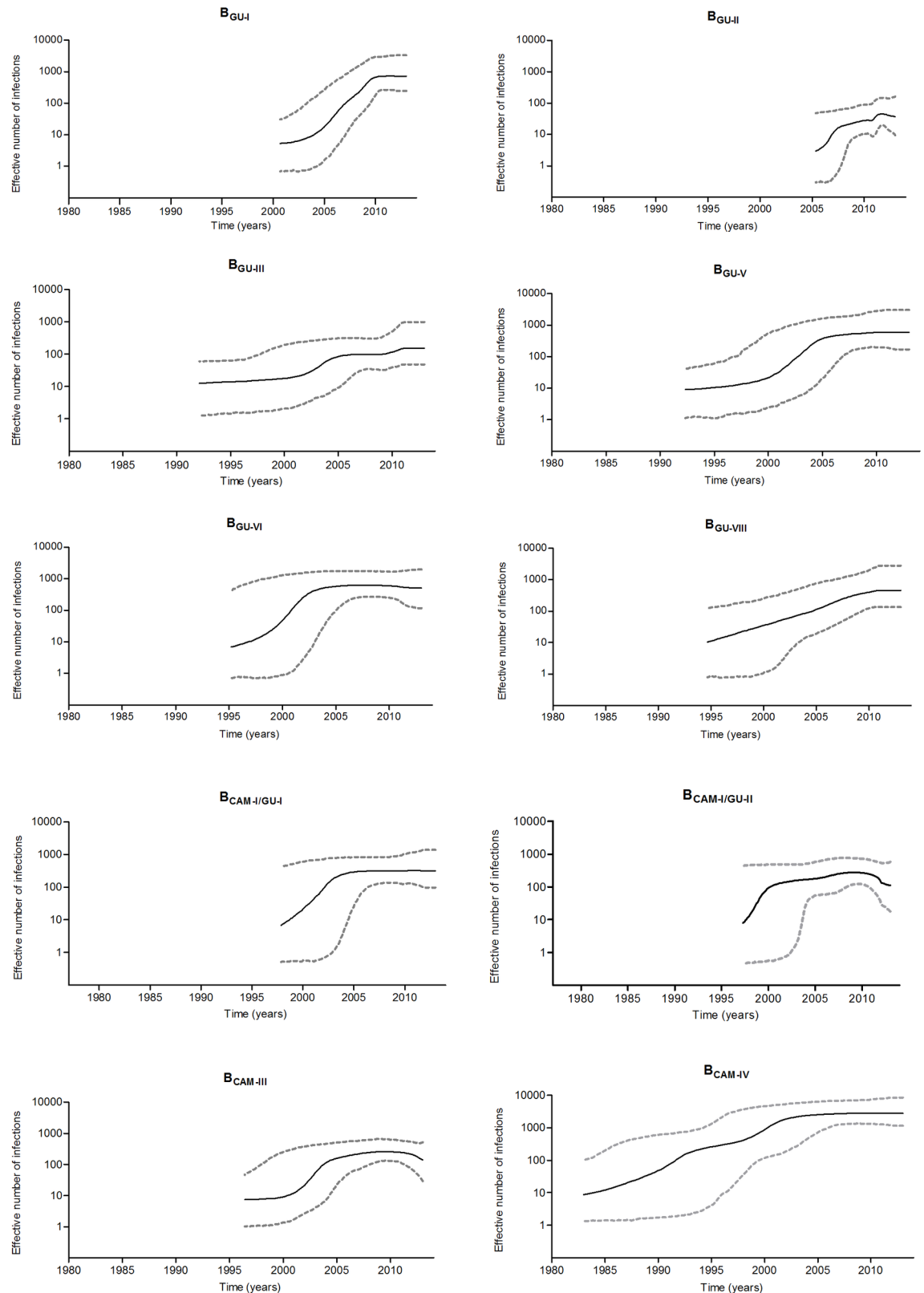
Fig 6. Demographic history of HIV-1 Guatemalan clades. Effective number of infections per year was estimated using a Bayesian skyline growth coalescent model for each one of the Guatemalan clades. Median estimates of the effective number of infections (solid line) and 95% High Posterior Density intervals of the estimates (dashed lines) are shown.

Table 3. Population dynamics estimates for Guatemalan HIV-1 subtype B clades.

Clade	Demographic model	T_{MRCA}[a]	Growth rate (yr^{-1})[a]
B_{GU-I}	Exponential	2000 (1995–2004)	0.58 (0.36–0.82)
	Logistic	2000 (1996–2005)	0.70 (0.39–1.11)
B_{GU-II}	Logistic	2005 (2002–2007)	3.59 (0.30–9.02)
	Exponential	2004 (2000–2007)	0.41 (0.19–0.66)
B_{GU-III}	Exponential	1993 (1983–2001)	0.24 (0.10–0.42)
	Logistic	1992 (1982–2001)	0.23 (0.00–2.34)
B_{GU-V}	Exponential	1993 (1985–2000)	0.33 (0.21–0.52)
	Logistic	1993 (1986–2000)	0.43 (0.21–0.76)
B_{GU-VI}	Logistic	1996 (1996–2001)	0.99 (0.48–1.72)
$B_{GU-VIII}$	Exponential	1995 (1987–2001)	0.36 (0.19–0.54)
	Logistic	1995 (1987–2001)	0.38 (0.13–0.68)
$B_{CAM-I/GU-A}$	Logistic	1998 (1993–2003)	1.46 (0.52–3.00)
$B_{CAM-I/GU-B}$	Logistic	1997 (1991–2002)	2.09 (0.61–5.03)
$B_{CAM-III/GU}$	Logistic	1997 (1990–2001)	0.68 (0.35–1.12)
$B_{CAM-IV/GU}$	Logistic	1984 (1975–1994)	0.40 (0.26–0.63)

[a]95% Higher Posterior Density intervals are shown in parentheses.

https://doi.org/10.1371/journal.pone.0203916.t003

among individuals infected by the same transmission route (heterosexual transmission), the mean estimated growth rate of these lineages varied considerably, ranging from 0.2 year^{-1} for clade B_{GU-III} to 2.1 year^{-1} for clade $B_{CAM-I/GU-II}$. This range is even higher than that estimated for country-specific $B_{PANDEMIC}$ clades circulating in the general population in different Latin American countries (0.2–0.9 year^{-1}) [7, 8, 36].

It is important to acknowledge some limitations of our study design. Even when the study sample size is large, sequences were obtained from a single HIV clinic. Additionally, the time frame for collecting HIV sequences was limited (2010–2013). Nevertheless, the Roosevelt Hospital is the largest of a total of 18 HIV clinics in the country, providing antiretroviral treatment for 30% of all registered HIV-infected persons in Guatemala [15], and the demographic characteristics (including female-to-male ratio, mean age, race, marital status, literacy) and geographical distribution of persons attending this clinic are highly representative of the whole population living with HIV in Guatemala [11, 16, 17]. Indeed, among the Roosevelt Hospital's clients, 45% reside outside the department of Guatemala [14, 16]. Our study group is highly representative of the Roosevelt Hospital cohort, as more than 80% of all adults diagnosed at this clinic during the enrolment period were included in the study (expecting approximately 450 total new cases per year [16]). Given these characteristics of the Roosevelt Hospital's cohort, we would not expect our main observations to vary significantly if additional participants enrolled in clinics from around the country were to be included.

Conclusions

In summary, our observations suggest that the Guatemalan HIV-1 subtype B epidemic is being driven by dissemination of multiple $B_{PANDEMIC}$ founder viral strains, some of them mostly restricted to Guatemala and others widely disseminated in the Central American region. Moreover, our results identify Guatemala and Honduras as hubs that have played a central role in dissemination of $B_{PANDEMIC}$ viral strains within Central America and further suggest the existence of different sub-epidemics with distinct evolutionary and demographic

dynamics within Guatemala and for which different targeted prevention efforts might be needed.

Supporting information

S1 Data. Selected HIV-1 subtype B PR-RT sequences. Fasta HIV-1 subtype B PR-RT Guatemalan sequences included in the study.
(TXT)

S1 Table. Origin and sampling interval of HIV-1 subtype B sequences.
(DOCX)

S2 Table. Distribution of HIV-1 subtype B sequences of Central American countries in Guatemalan B$_{CAM}$ clades.
(DOCX)

S3 Table. Best fit demographic model for the major HIV-1 Guatemalan clades. Log marginal likelihood (ML) estimates for the logistic (Log), exponential (Expo) and expansion (Expa) growth demographic models obtained using the path sampling (PS) and stepping-stone sampling (SS) methods. The Log Bayes factor (BF) is the difference of the Log ML between of alternative (H1) and null (H0) models (H1/H0). Log BFs > 3 indicates that model H1 is more strongly supported by the data than model H0.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Yaxelis Mendoza, Gonzalo Bello, Juan Miguel Pascale, Carlos Rodolfo Mejía-Villatoro, Gustavo Reyes-Terán.

Data curation: Yaxelis Mendoza, Claudia García-Morales.

Formal analysis: Yaxelis Mendoza, Claudia García-Morales, Gonzalo Bello, Daniela Garrido-Rodríguez, Santiago Avila-Ríos.

Funding acquisition: Santiago Avila-Ríos, Gustavo Reyes-Terán.

Investigation: Claudia García-Morales, Daniela Tapia-Trejo.

Methodology: Claudia García-Morales, Daniela Tapia-Trejo.

Project administration: Carlos Rodolfo Mejía-Villatoro, Santiago Avila-Ríos, Gustavo Reyes-Terán.

Resources: Amalia Carolina Girón-Callejas, Ricardo Mendizábal-Burastero, Ingrid Yessenia Escobar-Urias, Blanca Leticia García-González, Jessenia Sabrina Navas-Castillo, María Cristina Quintana-Galindo, Rodolfo Pinzón-Meza.

Supervision: Juan Miguel Pascale, Ingrid Yessenia Escobar-Urias, Carlos Rodolfo Mejía-Villatoro, Santiago Avila-Ríos, Gustavo Reyes-Terán.

Validation: Claudia García-Morales, Santiago Avila-Ríos.

Visualization: Yaxelis Mendoza.

Writing – original draft: Yaxelis Mendoza, Gonzalo Bello, Amalia Carolina Girón-Callejas, Ricardo Mendizábal-Burastero, Santiago Avila-Ríos.

Writing – review & editing: Yaxelis Mendoza, Gonzalo Bello, Amalia Carolina Girón-Callejas, Ricardo Mendizábal-Burastero, Santiago Avila-Ríos.

References

1. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(47):18566–70. Epub 2007/11/06. https://doi.org/10.1073/pnas.0705329104 PMID: 17978186; PubMed Central PMCID: PMC2141817.

2. Nadai Y, Eyzaguirre LM, Sill A, Cleghorn F, Nolte C, Charurat M, et al. HIV-1 epidemic in the Caribbean is dominated by subtype B. PloS one. 2009; 4(3):e4814. Epub 2009/03/13. https://doi.org/10.1371/journal.pone.0004814 PMID: 19279683; PubMed Central PMCID: PMC2652827.

3. Cabello M, Mendoza Y, Bello G. Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. PloS one. 2014; 9(8):e106045. Epub 2014/08/26. https://doi.org/10.1371/journal.pone.0106045 PMID: 25148215; PubMed Central PMCID: PMC4141835.

4. Paraskevis D, Pybus O, Magiorkinis G, Hatzakis A, Wensing AM, van de Vijver DA, et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology. 2009; 6:49. Epub 2009/05/22. https://doi.org/10.1186/1742-4690-6-49 PMID: 19457244; PubMed Central PMCID: PMC2717046.

5. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, Ashton-Prolla P, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. PloS one. 2011; 6(11):e27489. Epub 2011/12/02. https://doi.org/10.1371/journal.pone.0027489 PMID: 22132104; PubMed Central PMCID: PMC3223166.

6. Pagan I, Holguin A. Reconstructing the Timing and Dispersion Routes of HIV-1 Subtype B Epidemics in The Caribbean and Central America: A Phylogenetic Story. PloS one. 2013; 8(7):e69218. Epub 2013/07/23. https://doi.org/10.1371/journal.pone.0069218 PMID: 23874917.

7. Murillo W, Veras N, Prosperi M, Lorenzana de Rivera I, 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. Journal of virology. 2013; 87(13):7463–70. https://doi.org/10.1128/JVI.01602-12 PMID: 23616665

8. Mendoza Y, Martinez AA, Castillo Mewa J, Gonzalez C, Garcia-Morales C, Avila-Rios S, 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. 2014; 9(4):e95360. Epub 2014/04/22. https://doi.org/10.1371/journal.pone.0095360 PMID: 24748274; PubMed Central PMCID: PMC3991702.

9. UNAIDS. How AIDS changed everything—MDG6: 15 years, 15 lessons of hope from the AIDS response. Switzerland: Joint United Nations Programme on HIV/AIDS; 2015. p. ISBN: 978-92-9253-069-3.

10. Guatemala: evaluación final del plan estratégico nacional para la prevención, atención y control de ITS, VIH y SIDA, 2011–2015. Gobierno de Guatemala. Ministerio de Salud Pública y Asistencia Social.; 2015.

11. Garcia J. Vigilancia Epidemiológica del VIH, Guatemala 2017. Ministerio de Salud Pública y Asistencia Social, Departamento de Epidemiología. 2017. Available from: http://epidemiologia.mspas.gob.gt/files/Publicaciones2017/VIH/VIHaJunio2017JGrealizadaenagosto2017.pdf.

12. Informe Nacional sobre los Progresos Realizados en la Lucha Contra el VIH y sida. Ministerio de Salud Púlica y Asistencia Social. Programa Nacional de Prevención y Control de ITS/VIH/SIDA2014.

13. WHO. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection. Recommendations for a Public Health Approach. Second Edition. 2016. Available from: http://apps.who.int/iris/bitstream/handle/10665/208825/9789241549684_eng.pdf;jsessionid=04F13FC7759D72D44ED6AC6B2E5433F5?sequence=1.

14. Avila-Rios S, Garcia-Morales C, Garrido-Rodriguez D, Tapia-Trejo D, Giron-Callejas AC, Mendizabal-Burastero R, et al. HIV-1 drug resistance surveillance in antiretroviral treatment-naive individuals from a

reference hospital in Guatemala, 2010–2013. AIDS research and human retroviruses. 2015; 31 (4):401–11. Epub 2014/10/28. https://doi.org/10.1089/aid.2014.0057 PMID: 25347163.

15. MSPAS. Informe Nacional de la Cascada del Continuo de Atención en VIH, medición del indicador de sobrevida, Guatemala, 2015. Ministerio de Salud Pública y Asistencia Social de Guatemala. 2017. Available from: http://www.pasca.org/userfiles/InformeCascadadelContinuodeAtenciónenVIHGuatemala2015.pdf.

16. Giron-Callejas AC, Mendizabal-Burastero R, García L, Díaz S, Mejia-Villatoro C. Análisis de la vigilancia epidemiológica de VIH y Sida del Hospital Roosevelt, Guatemala 2002–2010. Guatemala: Hospital Roosevelt, Ministerio de Salud Pública y Asistencia Social. 2011. Available from: http://docplayer.es/13349096-Amalia-carolina-giron-callejas-trabajos-finales-nivel-intermedio.html.

17. Garcia JI, Samayoa B, Sabido M, Prieto LA, Nikiforov M, Pinzon R, et al. The MANGUA Project: A Population-Based HIV Cohort in Guatemala. AIDS Res Treat. 2015; 2015:372816. https://doi.org/10.1155/2015/372816 PMID: 26425365; PubMed Central PMCID: PMCPMC4575727.

18. 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 research. 1997; 25:4876–82. PMID: 9396791

19. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular biology and evolution. 2007; 24(8):1596–9. Epub 2007/05/10. https://doi.org/10.1093/molbev/msm092 PMID: 17488738.

20. 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 (19):3797–800. https://doi.org/10.1093/bioinformatics/bti607 PMID: 16076886.

21. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008; 25(7):1253–6. Epub 2008/04/10. doi: msn083 [pii] https://doi.org/10.1093/molbev/msn083 PMID: 18397919.

22. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52(5):696–704. PMID: 14530136.

23. 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(Web Server issue):W557–9. https://doi.org/10.1093/nar/gki352 PMID: 15980534.

24. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006; 55(4):539–52. Epub 2006/06/21. https://doi.org/10.1080/10635150600755453 PMID: 16785212.

25. Rambaut A. FigTree v1.4: Tree Figure Drawing Tool. Available from: http://treebioedacuk/software/figtree/. Accessed 2013 June 20. 2009.

26. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 2002; 161 (3):1307–20. PMID: 12136032.

27. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214. https://doi.org/10.1186/1471-2148-7-214 PMID: 17996036.

28. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics. 2009; 25 (11):1370–6. Epub 2009/04/17. https://doi.org/10.1093/bioinformatics/btp244 PMID: 19369496; PubMed Central PMCID: PMC2682525.

29. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4(5):e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862.

30. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009; 5(9):e1000520. Epub 2009/09/26. https://doi.org/10.1371/journal.pcbi.1000520 PMID: 19779555; PubMed Central PMCID: PMC2740835.

31. Rambaut A, Drummond A. Tracer v1.6. Available: http://treebioedacuk/software/tracer/. Accessed 2013 June 20. 2007.

32. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics. 2011; 27(20):2910–2. Epub 2011/09/14. https://doi.org/10.1093/bioinformatics/btr481 PMID: 21911333; PubMed Central PMCID: PMC3187652.

33. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005; 22(5):1185–92. https://doi.org/10.1093/molbev/msi103 PMID: 15703244.

34. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Molecular biology and evolution. 2012; 29(9):2157–67. Epub 2012/03/10. https://doi.org/10.1093/molbev/mss084 PMID: 22403239; PubMed Central PMCID: PMC3424409.

35. Crabtree-Ramirez B, Caro-Vega Y, Shepherd BE, Wehbe F, Cesar C, Cortes C, et al. Cross-sectional analysis of late HAART initiation in Latin America and the Caribbean: late testers and late presenters. PloS one. 2011; 6(5):e20272. Epub 2011/06/04. https://doi.org/10.1371/journal.pone.0020272 PMID: 21637802; PubMed Central PMCID: PMC3102699.

36. Mir D, Cabello M, Romero H, Bello G. Phylodynamics of major HIV-1 subtype B pandemic clades circulating in Latin America. AIDS. 2015; 29(14):1863–9. Epub 2015/09/16. https://doi.org/10.1097/QAD.0000000000000770 PMID: 26372392.

37. Cabello M, Junqueira DM, Bello G. Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. AIDS. 2015; 29(4):483–92. Epub 2015/01/30. https://doi.org/10.1097/QAD.0000000000000552 PMID: 25630042.

38. Morales-Miranda S, Alvarez-Rodriguez B, Arambu N, Aguilar J, Huamán B, Figueroa W, et al. Encuesta de Vigilancia de Comportamiento Sexual y Prevalencia del VIH e ITS, en poblaciones vulnerables y poblaciones clave. Guatemala: Universidad del Valle de Guatemala, 2013.

39. Arias J. Demografía de Guatemala. In: Lima FR, editor. Época contemporanea: de 1945 a la actualidad. Historia General de Guatemala. VI. Guatemala: Fundación para la Cultura y el Desarrollo; 1997. p. 195–213.

40. Aguilera G. La Guerra Interna, 1960–1994. In: Rojas F, editor. Época contemporanea: de 1945 a la actualidad. Historia General de Guatemala. VI. Guatemala: Fundación para la Cultura y el Desarrollo; 1997. p. 135–51.

41. Schroten H. La migración interna en Guatemala durante el período 1976–1981. Notas de Población. 1987.

42. Caballeros Á. Perfil migratorio de Guatemala 2012. 2013.

43. Leyva-Flores R, Aracena-Genao B, Serván-Mori E. Movilidad poblacional y VIH/sida en Centroamérica y México. Rev Panam Salud Publica. 2014; 36(6):143–9.

44. Montejo V. Los Refugiados Guatemaltecos en México, 1981–1992. In: Rojas F, editor. Época contemporanea: de 1945 a la actualidad. Historia General de Guatemala. VI. Guatemala: Fundación para la Cultura y el Desarrollo; 1997. p. 367–79.

45. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(12):4425–9. Epub 2005/03/16. https://doi.org/10.1073/pnas.0407534102 PMID: 15767575; PubMed Central PMCID: PMC555492.

46. 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(2):156–60. https://doi.org/10.1097/QAI.0b013e3181eb3002 PMID: 20703157

47. Chen JH-K, Wong K-H, Chan KC-W, To SW-C, Chen Z, Yam W-C. Phylodynamics of HIV-1 Subtype B among the Men-Having-Sex-with-Men (MSM) Population in Hong Kong. PloS one. 2011; 6 doi:(9): e25286. https://doi.org/10.1371/journal.pone.0025286.t001

48. Delatorre EO, Bello G. Phylodynamics of the HIV-1 Epidemic in Cuba. PloS one. 2013; 8(9):e72448. https://doi.org/10.1371/journal.pone.0072448 PMID: 24039765