Increase in human immunodeficiency virus 1 diversity and detection of various subtypes and recombinants in north-eastern Brazil

Kledoaldo Lima,[1,*] Élcio Leal,[2] Ana Maria Salustiano Cavalcanti,[3] Daniela Medeiros Salustiano,[3] Luzidalva Barbosa de Medeiros,[4] Sirleide Pereira da Silva[3] and Heloísa Ramos Lacerda[1]

Abstract

Purpose. Diverse human immunodeficiency virus 1 (HIV-1) subtypes and circulating recombinant forms are found in Brazil. The majority of HIV-1 molecular epidemiological studies in Brazil have been conducted in the southern and south-eastern regions of the country, although several recent studies in the north-eastern region have addressed this issue. The objective of this study was to molecularly characterize HIV-1 circulating in Pernambuco, north-eastern Brazil.

Methodology. A total of 64 samples were collected from 2002 to 2003, and another 103 were collected from 2007 to 2009. The protease and partial reverse transcriptase regions of the HIV-1 polymerase-encoding (*pol*) gene were sequenced, and subtyping, recombination and phylogenetic analyses were performed.

Results/Key findings. Subtype B (60.9 %) was found to be predominant, followed by HIV-1 F (31.4 %). Several BF recombinants (4.2 %), and BC and AG recombinants were also identified. The intra-subtype genetic diversity was estimated to be 0.065 (SD ±0.004) for HIV-1 B and 0.055 (SD ±0.004) for HIV-1 F, reflecting a greater accumulation of mutations in subtype B ($P<0.01$). More codons were found to be under positive selective pressure in samples collected from 2007 to 2009, from individuals with a T-cell count$\geq$200 cells mm^{-3} and from women. Coalescence data indicated that the subtype F population has been continuously expanding.

Conclusions. HIV-1 shows high genetic diversity in the state of Pernambuco. Thus, additional molecular evaluations of circulating strains will provide a better understanding of the epidemic and may lead to more effective preventive strategies.

INTRODUCTION

Human immunodeficiency virus 1 (HIV-1) shows high genetic diversity in Brazil. Subtype B is the most prevalent, followed by the subtypes F and C [1, 2]. Multiple circulating recombinant forms (CRFs) have been identified in Brazil, including CRF31_BC in the south [3]; CRF28_BF and CRF29_BF in Santos, São Paulo state [4]; CRF39_BF and CRF40_BF in Rio de Janeiro [5]; CRF_46BF in São Paulo [6]; CRF70_BF and CRF71_BF in Pernambuco [7]; and CRF72_BF in Minas Gerais [8]. In addition, multiple subtypes and minority recombinants have been identified, including subtypes A [9] and D [1, 10–13], and recombinants FD [11, 14] and FC [11]. There has been one report of autochthonous horizontal transmission of CRF02_AG in Rio de Janeiro [15]. Unique recombinant forms (URFs) of BF were identified early [16] in the Brazilian epidemic and subsequently disseminated to various regions of the country [17–20]. In addition, CRF45_cpx was recently detected in São Paulo [21].

The HIV-1 epidemic has been stable in most of Latin America since 2000, but stabilized in Brazil after 2004; two-thirds of the people living with HIV in South and Central America

reside in Brazil (www.unaids.org/en/dataanalysis/kno-wyourepidemic/epidemiologypublications/). In Brazil, 656 701 AIDS cases have been reported, with 13.5 % (88 830) of these being in the north-eastern region. The state of Pernambuco has the highest incidence of AIDS in the country, and the city of Recife has the second highest incidence of AIDS among the capital cities in the region. The epidemic is most prevalent in high-risk populations, such as drug users, men who have sex with men (MSM) and sex workers [22].

HIV-1 strains in north-eastern Brazil show high molecular diversity, with the subtypes B, F, C and D circulating [13, 23–25]. A high frequency of BF recombinants has been reported in the state of Bahia [13, 26], and subtype F, CRF70, and CRF71_BF are common in the state of Pernambuco [7, 17].

Co-circulation of multiple subtypes may facilitate the generation and dispersal of recombinant viruses [4, 27]. Recombination among subtypes B and F has been occurring since the introduction of the latter subtype in South America [28, 29]. BF recombinants have played an important role in the HIV-1 epidemic in South America, dispersing rapidly after their emergence [27, 30]. Identification and characterization of recombinant viruses is important, because recombination can have a major impact on HIV-1 evolution, transmission vector specificities, virulence, pathogenesis, and potential for host immune evasion and *in vitro* fitness recovery [31, 32].

Since 2002, the incidence of AIDS has markedly increased in northern and north-eastern Brazil, which represents some of the country's poorest areas (ftp://ftp.ibge.gov.br/Contas_Regionais/2010/pdf/tab02.pdf) [22]. The current study aimed to characterize the molecular epidemiology of HIV-1 in the state of Pernambuco in north-eastern Brazil by identifying the types and frequencies of viral variants in patients attending a referral hospital for infectious diseases and at some voluntary counselling and testing centres (VCTs).

METHODS

Study population

Samples were obtained from patients attending the Hospital of the Federal University of Pernambuco, Recife, Brazil, between 2002 and 2003, in a study by de Medeiros *et al.* [25], and from individuals who sought HIV diagnosis at the five largest VCTs in Pernambuco, between 2007 and 2009, in a study by Cavalcanti *et al.* [17]. There was no overlap in patients between the two study populations. All samples were collected in Pernambuco state, north-eastern Brazil, and all participants provided written informed consent. Only samples from treatment-naive individuals were included. Samples codes, dates of sampling, viral subtypes and GenBank numbers are presented in Table S1 (available in the online Supplementary Material). Convenience sampling was employed, and samples were collected from participants representing different groups: heterosexual (HTS)

individuals, MSM, sex workers, pregnant women, injection drug users and blood transfusion recipients. All data were obtained from the databases of the above studies and were anonymized [17, 25]. Additional sociodemographic and behavioural data were collected from medical records at the institutions where the participants were first diagnosed.

CD4$^+$ T-lymphocyte counts and viral loads that were not available in research databases were obtained from the records of the Central Public Health Laboratory of Pernambuco (LACEN-PE); results collected closest to the date of HIV-1 diagnosis were used. CD4$^+$ T-cell counts were measured by flow cytometry (FACSCalibur; Becton Dickinson) and viral loads were measured by a nucleic acid sequence-based amplification assay (NASBA; Organon Teknika) and branched-chain DNA assay (VERSANT HIV-1 RNA 3.0 assay; Bayer). BED-capture enzyme immunoassay results, used to determine recent or long-term HIV-1 infection, were obtained in a study by Cavalcanti *et al.* [17].

The state of Pernambuco is located in north-eastern Brazil. It borders the Atlantic Ocean and has a population of 9.2 million people in an area of 98 148 km^2 (www.ibge.gov.br/estadosat/perfil.php?sigla=pe). This study was approved by the Ethics Committee of the Health Sciences Center of the Federal University of Pernambuco (CCS-UFPE), protocol no. 114 722, and the funding source was the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Ministry of Education, Brazil.

Sequencing of the HIV-1 polymerase-encoding (*pol*) gene

Initially, viral RNA was extracted from plasma, using a QIAamp RNA mini kit (Qiagen). Sequencing of the entire protease (PR) and partial reverse transcriptase (RT) of *pol* was performed using the ViroSeq HIV-1 genotyping system (Abbott Laboratories) [25] or the TRUGENE HIV-1 genotyping assay (Siemens Diagnostics) [17], and an ABI PRISM 3100 DNA automated sequencer (Applied Biosystems) and OpenGene sequencing system (Siemens Diagnostics), respectively. All procedures were performed according to the manufacturer's instructions. Fragments of 918 bp were obtained that encompassed the PR (positions 2262–2549) and partial RT (positions 2661–3290; position numbers based on a comparison with strain HXB2, GenBank accession number K03455). The GenBank accession numbers for the 167 HIV-1 nucleotide sequences are KX020958–KX021124.

Phylogenetic analysis

The HIV-1 *pol* sequences were initially analysed using the REGA automated tool for HIV-1 and -2 subtyping (version 2.0) (www.bioafrica.net/rega-genotype/html/subtypinghiv.html) to determine viral subtypes. An additional analysis of each sequence was conducted using BLAST (www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html), and similar sequences were obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov/content/index) for comparison.
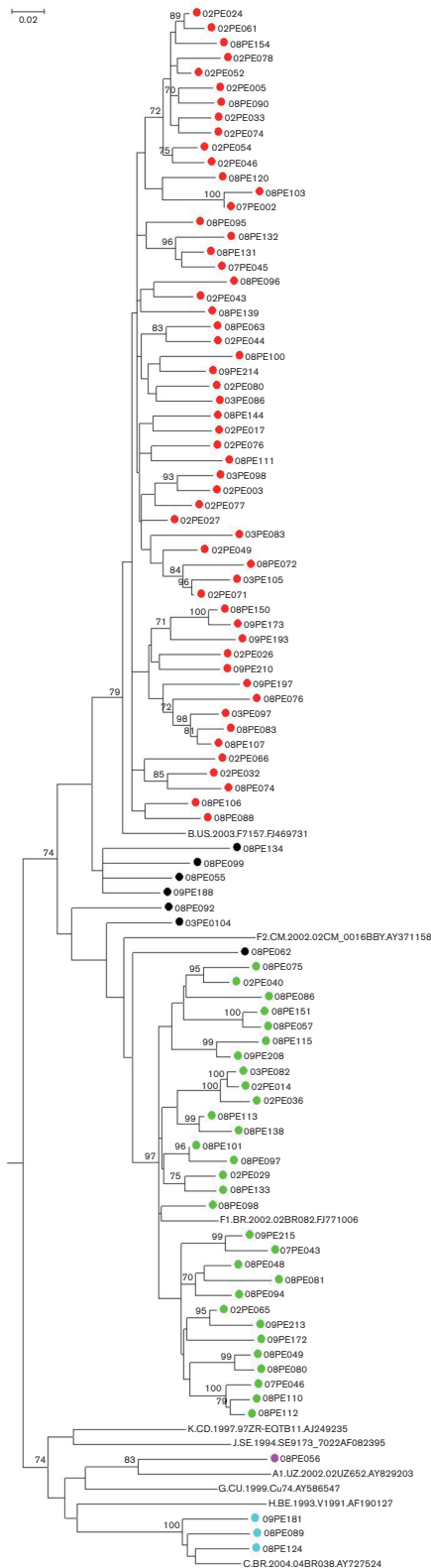
Fig. 1. Phylogenetic inference of HIV-1 in Pernambuco, north-eastern Brazil. The alignment contained 891 bp comprising some query and reference sequences; resistance mutations were excluded. The tree structure was determined by ML analysis implemented in MEGA 5 with 1000 bootstrap replicates. Tree nodes with a bootstrap ≥70 are demonstrated. Red circles, HIV-1 B; green circles, HIV-1 F; black circles, BF recombinants; blue circles, HIV-1 C; purple circle, HIV-1 H. Phylogenetic inference with all query sequences is available in Fig. S1.

Multiple alignments were performed using CLUSTALX [33], followed by manual editing using BioEdit software [34]. Sites with gaps or ambiguities were excluded, and the reading frame of the *pol* gene was found to be conserved.

The nucleotide substitution model was selected following a likelihood ratio test with Bayesian information criterion, implemented in MEGA 5.0 [35]. The general time-reversible model with gamma distribution and proportion of invariable sites (GTR+G+I) was selected for implementation. Phylogenetic inferences were made using maximum likelihood (ML) criteria in MEGA 5.0. Statistical support for tree topologies was determined by a bootstrap analysis with 1000 replicates. Phylogenetic networks, assembled in SplitsTree4 [36] using the NeighborNet method, were used to confirm the phylogeny. Trees were visualized in FigTree version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Screening for HIV-1 intersubtype recombination

Initially, we used the BLAST genotyping tool (www.ncbi.nlm. nih.gov/projects/genotyping/formpage.cgi) to identify subtypes and recombinants from viral sequences. Each query sequence was compared, by BLAST, with reference sequences in the Human Immunodeficiency Virus 1 (HIV-1) 2009 RefSeq database.

To determine the extent of recombination in HIV sequences, we used SIMPLOT software version 3.5.1 [37] and RDP software v.4 [38]. The following parameters were used in SIMPLOT: F84 nucleotide substitution model with 1000 bootstrap replicates, sliding window of 200, 20 bp steps and 90 % confidence. The ratio of transitions/transversions was determined empirically for each alignment evaluated. We applied the RDP package, using a variety of recombination detection methods, including 3Seq, which performs a nonparametric test to detect recombination in triplets of sequences [39], and maximum χ^2, a method developed by Maynard-Smith [40] that uses variable/ invariable sites to identify recombination in pairs of sequences. This method generates random sequence pairs, and the significance level is determined by the proportion of simulated

Table 1. Frequencies of HIV-1 B and F in 2002–2003 and 2007–2009 in Pernambuco, north-eastern Brazil

We took into consideration only the prevalent cases of HIV-1 B and F, not including other non-B subtypes in the analysis. Chi-square test: *P*=0.01.

Year	HIV-1 subtype	
	B [*N* (%)]	F [*N* (%)]
2002–2003	48 (77.4)	14 (22.6)
2007–2009	55 (58.5)	39 (41.5)

Table 2. Breakpoint positions of B and F genomic segments in the *pol* gene fragment of HIV-1 BF recombinants

Sequence ID	Breakpoint position* (parental subtype)	RDP method
08PE055	2262–2628 (F); 2629–3290 (B)	GeneConv
08PE062	2262–3054 (F); 3055–3290 (B)	GeneConv
08PE092	2262–2632 (F); 2633–3290 (B)	MaxChi
08PE099	2262–2644 (F); 2645–3290 (B)	GeneConv
03PE104	2262–2976 (F); 2977–3290 (B)	GeneConv
08PE134	2262–2642 (F); 2643–3290 (B)	GeneConv
09PE188	2262–2640 (F); 2641–3290 (B)	GenConv

MaxChi, maximum χ^2.

*Positions are numbered according to the *pol* gene of the HXB2 reference sequence (GenBank accession no. K03455).

sequence pairs with maximum χ^2 values higher than those in the real data. The maximum match χ^2 (chimaera) is a modification of Smith's method; It uses variable sites to calculate maximum χ^2 match statistics [41]. In addition, GeneConv was used to detect gene conversions (recombination) by evaluating conserved substitutions in two sequences [42]. Although evolutionary methods are not explicitly implemented in Gene-Conv, it offers a robust recombination detection method with a low number of false positives, including events resulting from rate heterogeneity and natural selection [43]. Initially, we used the default parameters; then, the parameters were

optimized to reduce the number of false positives. In addition, window sizes of 20 to 50, as well as a Bonferroni correction with *P* values of 0.05 and 0.001, were utilized. We also used INTERVAL [44] in the RDP package, which uses an approximate likelihood approach with a Bayesian reversible-jump Markov chain Monte Carlo scheme to estimate the alignment-wide population-scaled recombination rate (ρ). Here, we used a mutation rate of 0.003 per generation per site and population sizes of 10^2 to 10^3 (previously calculated in BEAST).

Coalescent analysis

We used a Bayesian Markov chain Monte Carlo (BMCMC) coalescent framework to estimate ancestral genealogy, phylodynamics and evolutionary parameters such as nucleotide substitutions per year and time to the most recent common ancestor (tMRCA). The BMCMC method incorporates the uncertainty of measurements by considering the errors intrinsic to both tree reconstruction and the coalescent method. The uncertainty is estimated by the highest posterior density interval, which contains 95 % of the marginal posterior distribution. After the Markov chain has run long enough so that the parameter space has been sufficiently sampled, which is indicated by an effective sample size (ESS) greater than 100, and the parameters become stable, the mean and 95 % highest posterior density intervals of each parameter can be calculated. The Hasegawa-Kishino-Yano model plus a gamma correction (Γ) was applied to all

Fig. 2. Bayesian coalescent inference of population size history of subtype B from Pernambuco, north-eastern Brazil. Fragments of 918 bp were obtained encompassing the PR and part of the RT, corresponding to positions 2262–2549 from the PR and 2661–3290 from the RT, relative to a reference strain HXB2 (GenBank accession no. K03455). Reference sequences were obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov/content/index).
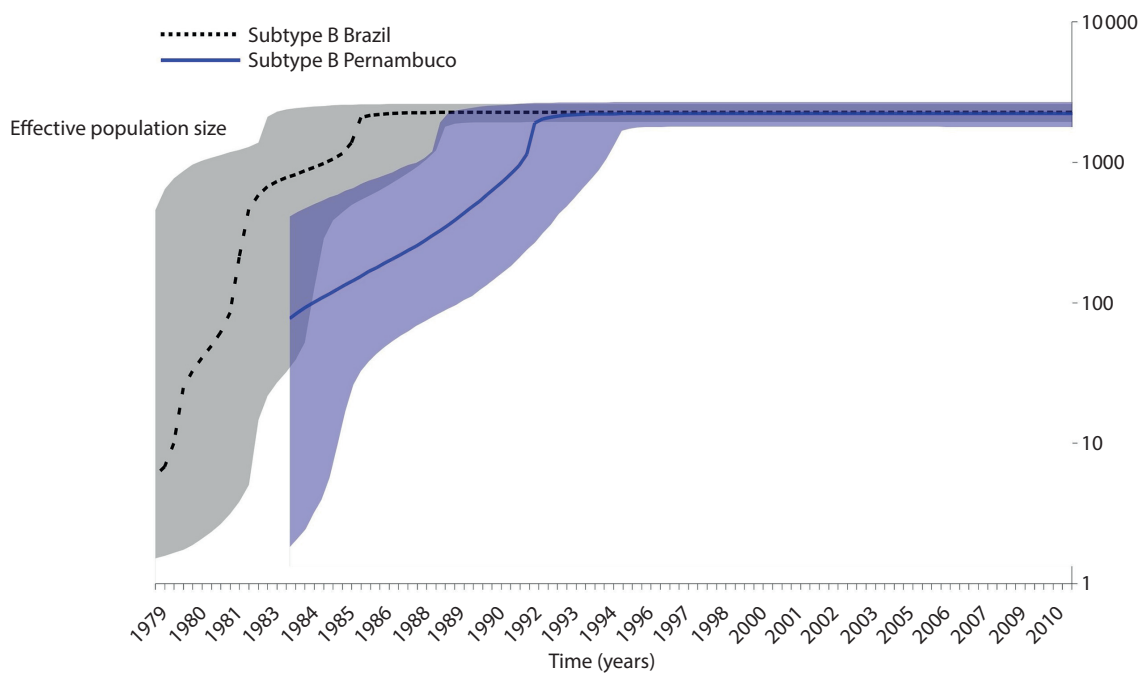
Fig. 3. Bayesian coalescent inference of population size history of subtype F from Pernambuco, north-eastern Brazil. Fragments of 918 bp were obtained encompassing the PR and part of the RT, corresponding to positions 2262–2549 from the PR and 2661–3290 from the RT, relative to a reference strain HXB2 (GenBank accession no. K03455). Reference sequences were obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov/content/index).
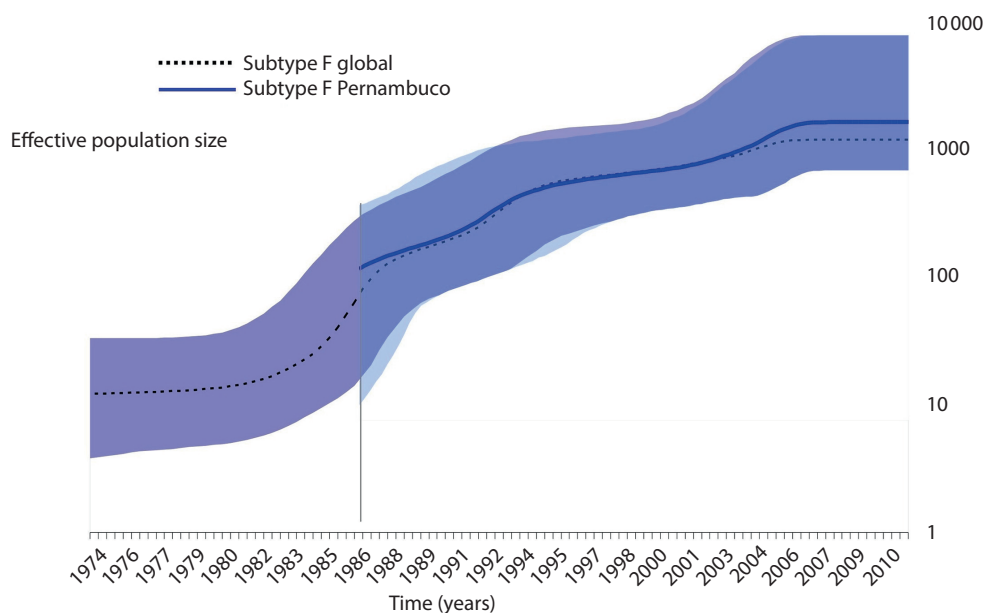
analyses, and the evolutionary and demographic parameters were iteratively adjusted.

The Bayesian skyline plot method provides an unbiased estimate of genetic diversity over time [expressed as the product of the ESS (Ne) and generation time (τ)] because a wide range of demographic scenarios are considered, without formally assuming any determined model [45]. Since we used absolute time (years) to scale branch lengths and did not assume a specific generation time, our estimates of $Ne.\tau$ reflect only the relative genetic diversity of viral strains over time. Besides, the Bayesian skyline plot was the best-fit model when compared with the constant coalescent model. Sequences sampled at different times (heterochronous samples) were used to estimate the ancestral genealogy, incorporating evolutionary parameters such as nucleotide substitution rate per year and growth rate over time. MCMC processes ran for 3×10^8 generations with the initial 10 % of each run discarded as burn-in. The convergence of chains was evaluated using Tracer 1.5 (available at http://beast.bio.ed.ac.uk/), and runs were accepted when all parameters presented an ESS greater than 200. Two independent chains were run for each dataset and combined using LogCombiner software [45, 46]. All analyses were performed with BMCMC approaches implemented in BEAST version 1.8.1 [46]. To estimate diversity over time ($Ne.\tau$), the mutation rate was fixed at 1.5×10^{-3}, and prior probability distributions were assumed, i.e. a normal distribution for a mean sample size of 400±150 and for a mean clock rate of 1.0×10^{-3}±0.001. Parameters were fixed to avoid a branch

Table 3. Codons under positive selective pressure in the HIV-1 B and F *pol* gene

Variable	Subtype	Codon*
T-cell counts†		
<200	B (*n*=32)	PR: –
		RT: –
	F (*n*=11)	PR: –
		RT: 173
≥200	B (*n*=49)	PR: 37, 77, 93
		RT: 161, 199, <u>210</u>‡
	F (*n*=29)	PR: 12, 19, 62
		RT: 172, 199
Exposure category		
MSM	B (*n*=33)	PR: –
		RT: 199, 244
	F (*n*=10)	PR: –
		RT: –
HTS female	B (*n*=36)	PR: 37
		RT: 209, 243
	F (*n*=25)	PR: 12, 63
		RT: 173, 200
HTS male	B (*n*=31)	PR: 37
		RT: –
	F (*n*=17)	PR: –
		RT: –

*Sequences with 918 bp (308 codons).
†Cells mm^{-3}.
‡Drug resistance codon.

length bias caused by a uniform prior distribution [47], and new estimates for sample size and clock rate were obtained. Next, to calibrate the tMRCA, a prior for the mutation rate of 1.0×10^{-3} was set. Fixed values (obtained at the first stage of the analysis) of $1.0 \times 10^{-3} \pm 0.001$ for the clock rate and 1200 ± 150 for the sample size were also used. These conditions allowed us to re-estimate the diversity ($Ne.\tau$) and date of nodes of the HIV-1 strains. In addition to the sequences generated in this study, we also analysed 141 HIV-1 B sequences from Brazil and 13 HIV-1 F sequences obtained worldwide to infer population sizes and the tMRCA (Supplementary Material S1).

Analysis of genetic diversity and positive selective pressure in HIV-1 *pol* subtypes B and F

A genetic diversity analysis was performed to compare the pairwise distances between sequences collected at different times. We determined the genetic distances among sequences belonging to subtypes B and F collected in 2002–2003 and 2007–2009. The intra-subtype genetic diversity among samples collected in each time period was inferred by estimating the nucleotide diversity (π) in MEGA 5.0.

To estimate the synonymous and nonsynonymous substitution rates across amino acid sites in the HIV-1 sequences, we used a random effects likelihood (REL) approach [48, 49]. REL is a phylogeny-based framework that explicitly allows both the synonymous and non-synonymous substitution rates to vary from site to site, and it uses a general underlying nucleotide substitution model to independently calculate the dS and dN rates. The REL method was described in Hyphy version 2.2 [50] and implemented in MEGA 5.0. Sequences were analysed by subtype (B and F), CD4 cell counts and type of exposure.

Statistical analysis

The chi-square test was performed to compare the proportions of subtypes B and F in each period (period 1, 2002–2003; period 2, 2007–2009). Statistical significance was set at 5 % ($P<0.05$). Age was given as the mean±SD. CD4$^+$ cell counts and viral loads were given as medians and interquartile intervals. Data were analysed using Stata V.13 software (Stata).

RESULTS

Characteristics of the study population

No participants received drug therapy during the study. The mean age of the patients was 34 years, and the majority were male (58.1 %). Regarding women, 13 % were pregnant. The main exposure categories were HTS (29.9 % were male HTS and 41.9 % were female HTS) and MSM (28.1 %), with a low prevalence of injection drug users and non-injection drug users (1.8 %). Additionally, 6.6 % of individuals reported sexual relationships with HIV-positive partners. The median CD4$^+$ cell count and viral load were 210 cells mm^{-3} and 14 731 copies ml^{-1}, respectively.

Molecular genotyping

There was a high frequency of subtypes B ($n=103$, 61.6 %) and F ($n=53$, 31.7 %), and the presence of HIV-1 C ($n=3$, 1.8 %), BF recombinants ($n=7$, 4.2 %) (Fig. 1) and one AG recombinant (Fig. S1). A ML phylogenetic analysis of all sequences is shown in Fig. S2. One HIV-1 C strain was classified as a BC recombinant after a recombination analysis (Fig. S3). The reconstructed phylogenetic networks confirmed the ML analysis (Fig. S4). Both methods agreed with the classification of HIV-1 subtypes and BF recombinants. Networks that contained recombinants were more reticulated than those without.

Among subtype F strains, 49 % (26/53) were from HTS males, 32 % (17/53) were from HTS females and 19 % (10/53) were from MSM, although there was no statistically significant association between exposure category and subtype F in Pernambuco ($P=0.061$). Using only the frequencies of subtypes B and F, which comprised more than 90 % of circulating HIV-1 strains in the region, a statistically significant increase in the frequency of subtype F was observed between the two time points (i.e. 2002–2003 and 2007–2009) (Table 1).

We found a low prevalence of subtype C (2/168, 1.2 %). Sample 09PE181, obtained in 2009, was a BC recombinant (Fig. S3). Epidemiological data and BED-capture enzyme immunoassay revealed that 09PE181 was a recent infection in a man with a male HIV-positive sexual partner. Recombinant AG (08PE056) was detected in a pregnant woman with a long-term infection.

Recombination analysis

Initially, we used default parameters to screen for recombination signals in the alignment. Screening correctly identified all BF intersubtype sequences that had been previously characterized by phylogenetic analysis. To determine breakpoint positions, the following references were used: subtype A, 02UZ652.AY829203; subtype B, US.2003.F7157. FJ469731; subtype C, 04BR038.AY727524; subtype F, 02BR082.FJ771006; subtype H, BE.1993.VI991.AF190127; subtype K, CD.1997.97ZR-EQTB11.AJ249235; subtype J, SE.1994.SE9173_7022.AF082395; and subtype G, CU.1999. Cu74.AY586547). To locate the breakpoints, we chose the method that resulted in the highest probability of recombination for each intersubtype sequence (Table 2). Interestingly, some BF URFs (08PE055, 08PE092, 08PE099, 08PE134 and 09PE188) had recombination breakpoints similar to those of CRFs 28 and 29 BF in the genomic region analysed. Recombination rates (ρ) were estimated from the alignment containing all 103 subtype B (0.8 ± 10^{-4}), 53 subtype F (1.2 ± 10^{-4}) and 7 intersubtype BF sequences (1.5 ± 10^{-4}).

Population dynamics of HIV-1 in Pernambuco, Brazil

We analysed variations in the population sizes of subtypes B and F (Figs 2 and 3). A Bayesian coalescent inference

method was applied to compare historical population sizes of subtype B in Pernambuco with those of Brazilian samples from the south-eastern region (mainly isolates from São Paulo, Rio de Janeiro and Minas Gerais) isolated from 1990 to 2010 (Fig. 2). These patterns of population growth and decline were quite similar between subtype B in Brazil and in Pernambuco, with steep exponential growth followed by a plateau in diversity. The main difference was the timing of infection, with the introduction of HIV-1 B in Pernambuco occurring during the exponential phase of growth in HIV-1 infections in the south-east. The population dynamics of subtype F were less clear, owing to the limited number of sequences available. Here, we compared the diversity of samples from Pernambuco with that of global subtype F reference strains (Supplementary Material S1). The HIV-1 population in Pernambuco showed a moderate growth phase, followed by a plateau in diversity, and was introduced later than the emergence of global subtype F strains (Fig. 3). The coalescent method was also applied to estimate the date that HIV-1 was introduced into Pernambuco, with 1978 inferred as the tMRCA [95 % confidence interval (CI) 1971–1981] and 1982 (95 % CI 1977–1986), respectively, for subtypes B and F.

Increase in intrasubtype diversity of subtypes B and F

The mean genetic distances for subtype B strains (years 2002–2003 and 2007–2009) were 0.056 (SD ±0.004) and 0.072 (SD ±0.004), respectively, whereas for subtype F, the distances were 0.045 (SD ±0.004) and 0.064 (SD ±0.004), respectively. In both subtypes, diversity increased between the two sampling periods ($P<0.0000001$). The overall intrasubtype genetic diversity was estimated to be 0.065 (SD ±0.004) for HIV-1 B and 0.055 (SD ±0.004) for HIV-1 F, reflecting a greater accumulation of mutations in subtype B ($P<0.0000001$).

Positive selective pressure on HIV-1 B and F *pol*

Codons identified as under positive selection in subtypes B and F strains were examined in association with T-cell count and type of exposure (Table 3). Codon positions were numbered according to the HXB2 reference sequence (GenBank accession no. K03455).

In subtype B, codon 37 (PR) showed the highest frequency of positive selection; for HIV-1 F, the codons most frequently under positive selection were 12 (PR) and 173 (RT). A higher number of codons under positive selection was observed in samples collected from individuals with a T-cell count ≥ 200 cells mm^{-3}. HTS females presented a higher number of codons under positive selection than HTS males.

DISCUSSION

To characterize HIV-1 strains circulating in Pernambuco, a state in north-eastern Brazil, 167 sequences of the PR-encoding gene and the initial two-thirds of the RT-encoding gene were analysed. The cross-sectional design and convenience sampling used in this study presented some limitations to determining the characteristics of the epidemic. Although the phylogenetic analyses were conducted on a small genomic *pol* fragment, some authors have reported advantages to using the *pol* region to evaluate inter-subtype recombination because it is an important recombination hot spot [18, 51]. We found high viral diversity in our sample dataset: subtypes B (61.7 %), F (31.7 %), C (1.2 %), BF (4.2 %), and recombinants AG and BC were detected.

The high frequency of HIV-1 subtype B (61.7 %) is consistent with it being the first subtype isolated in the Western world and its subsequent spread worldwide [52, 53]. However, some studies have shown that, in Pernambuco, there is a lower frequency of subtype B than in other states in north-eastern Brazil, such as Bahia and Ceará, where the frequency ranges from 77 to 89 % and 81 to 85 %, respectively [13, 26, 54–57].

Recently, Alencar *et al.* [58] showed a high proportion of subtype F strains circulating in Pernambuco (24 %) by analysing the *pol* region, which is consistent with our data. However, Pessôa *et al.* [21] evaluated the complete genomes of the HIV-1 subtype F strains sequenced by Alencar *et al.* [58] and showed that, among 24 samples previously determined to be HIV-1 F using the *pol* region, 23 were BF recombinants, with 4 CRF70_BF1 and 11 CRF71_BF1 novel recombinant types.

Although some studies indicate the dissemination of HIV-1 BF in South America [27, 59, 60], the frequency of these recombinants in Brazil varies by geographical region and even local area, indicating that factors other than proximity influence their dispersion [61]. In north-eastern Brazil, a high percentage (approximately 20 %) of BF recombinants among all analysed strains was reported in Bahia [13, 26, 62], whereas in Ceará it was less than 2.7 % [26, 56]. The BF URFs detected in samples collected in 2002–2003 indicated that these strains have been circulating in the state of Pernambuco for a long time. The movement of these recombinants in Brazil was evident in samples collected in the early to mid-1990s [55].

We reanalysed the results obtained by the ML and NeighborNet methods to avoid a bias in the phylogenetic data due to the presence of BF recombinants. In the network analysis, BF recombinants were located between clades B and F, and networks with recombinants were more reticulated than those without BF sequences. Thus, different phylogenetic inference methods corroborated the original identification of recombinants and other HIV-1 subtypes.

A low prevalence of HIV-1 subtype C (*n*=2, 1.2 %) was found, concordant with other reports [23, 54]. A recombinant BC strain was detected, to our knowledge, for the first time in the state of Pernambuco and for the second time in north-eastern Brazil, with the first reported in Salvador, Bahia [57]. The low prevalence of HIV-1 C in the north-eastern region contrasts with its high frequency in southern

Brazil; prevalence rates up to 63.9 % have been reported in Santa Catarina [63, 64]. Despite the low frequency in some parts of Brazil, several authors have shown its expansion to other regions [65–67].

To our knowledge, this is the first time that an AG recombinant has been detected in north-eastern Brazil. The strain was isolated in 2008 from a pregnant woman with a long-term infection. AG recombinants have been reported in other Brazilian regions: CRF02_AG was reported in Belém (northern region) [12], and several studies have found AG recombinants in Rio de Janeiro (south-eastern region). These studies include reports [15, 68, 69] showing multiple routes of transmissions and the gradual spread of AG recombinants.

Coalescent inferences of population size indicated that the introduction of subtype B occurred earlier in other regions of Brazil (the south-east) than in Pernambuco, corroborating previous data showing that HIV-1 was introduced into the country in Rio de Janeiro and São Paulo [70]. In addition, there was a significant increase in the frequency of HIV-1 F strains. These findings support the hypothesis that HIV-1 diversity is increasing in Brazil, with the spread of non-B subtypes, CRFs and URFs [71–75]. The subtype F population is expanding around the world and in Brazil. In addition, the introduction of subtype F in Pernambuco occurred later than in other parts of the country (1982, 95 % CI 1977–1986), while Bello *et al.* [76] estimated that this introduction in Brazil occurred in 1977 (95 % CI 1963–1988). Several hypotheses have been presented to explain the expansion of HIV-1 non-B subtypes in Pernambuco, such as the great economic growth in the state in the last decade (2001–2010) that resulted from industrial activity in the largest port in north-eastern Brazil, and the construction of refineries and petro-chemical plants that may have affected population immigration (www.worldbank.org/pt/news/press-release/2013/06/25/wb-brazil-pernambucos-new-inclusive-growth-program-will-benefit-nine-million-residents). In addition, the metropolitan area of Recife (the state capital) has one of the main industrial plants and the largest service centre in the north-east, making it a hub for technology, medicine, academia and research. In addition, the state is a hub for tourism and culture (http://www2.recife.pe.gov.br/pagina/metropole-das-oportunidades). Another factor that may have contributed to the increased incidence of non-B subtypes observed in the second sampling period (2007–2009) is a decrease in the frequency of MSM with the 'heterosexualization' of the HIV epidemic in Brazil [22]. The particular characteristics of the epidemic in this region may have repercussions on antiretroviral treatment strategies, disease progression, vaccine development and laboratory monitoring quality. Therefore, it is critical to continue characterizing non-B subtypes and examining their virus–host interactions.

Among the codons under positive selection, only codon 210 of RT has been directly associated with antiretroviral resistance. Codons under positive selective pressure were more frequent in individuals with T-cell counts >200 cells mm^{-3}

and in HTS women. Leal *et al.* [77] showed that codons associated with antiretroviral therapies under positive selection could be detected in drug-naïve individuals; however, the intensity of positive selection is higher at these codons in individuals receiving antiretroviral therapy. Although Arenas [78] revealed that some PR inhibitors promoted adaptation (dN/dS), our data demonstrated that PR had fewer codons under selective pressure than RT. Two explanations are possible. First, when we obtained the samples for sequencing, nucleoside and non-nucleoside RT inhibitors were commonly used as first-line therapeutics. Second, when PR inhibitors were introduced as the first-line therapeutic regimen, lopinavir/ritonavir were the drugs used and these drugs showed little effect on the global dN/dS according to Arenas [78].

In summary, our study characterized HIV-1 strains between 2002 and 2009, and found an increase in circulating BF intersubtypes and nucleotide and codon diversity during the study period. These increases coincide with an increasing number of HIV cases detected in the region.

References
1. Santos AF, Sousa TM, Soares EA, Sanabani S, Martinez AM *et al.* Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *Aids* 2006;20: 2011–2019.

2. de Souza AC, de Oliveira CM, Rodrigues CL, Silva SA, Levi JE. *Short communication:* molecular characterization of HIV type 1 BF *pol* recombinants from São Paulo, Brazil. *AIDS Res Hum Retroviruses* 2008;24:1521–1525.

3. Dias CF, Nunes CC, Freitas IO, Lamego IS, Oliveira IM *et al.* High prevalence and association of HIV-1 non-B subtype with specific sexual transmission risk among antiretroviral naïve patients in Porto Alegre, RS, Brazil. *Rev Inst Med Trop Sao Paulo* 2009;51: 191–196.

4. Filho DJDS, Sucupira MC, Casiero MM, Sabino EC, Diaz RS *et al.* Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res Hum Retroviruses* 2006;22:1–13.

5. Guimarães ML, Eyer-Silva WA, Couto-Fernandez JC, Morgado MG. Identification of two new CRF_BF in Rio de Janeiro State, Brazil. *Aids* 2008;22:433–435.

6. Sanabani SS, Pastena ER, Neto WK, Martinez VP, Sabino EC. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. *Virol J* 2010;7:74.

7. Pessôa R, Watanabe JT, Calabria P, Felix AC, Loureiro P *et al.* Deep sequencing of HIV-1 near full-length proviral genomes identifies high rates of BF1 recombinants including two novel circulating recombinant forms (CRF) 70_BF1 and a disseminating 71_BF1 among blood donors in Pernambuco, Brazil. *PLoS ONE* 2014;9: e112674.

8. Pessôa R, Carneiro Proietti AB, Busch MP, Sanabani SS. Identification of a novel HIV-1 circulating recombinant form (CRF72_BF1) in deep sequencing data from blood donors in southeastern Brazil. *Genome Announc* 2014;2:e00386-14.

9. de Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH *et al.* Prevalence, estimated HIV-1 incidence and viral diversity among people seeking voluntary counseling and testing services in Rio de Janeiro, Brazil. *BMC Infect Dis* 2010;10:224.

10. Couto-Fernandez JC, Eyer-Silva WA, Guimarães ML, Chequer-Fernandez SL, Grinsztejn B *et al.* Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil. *AIDS Res Hum Retroviruses* 2006;22:207–211.

11. Ferreira AS, Cardoso LP, Stefani MM. Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. *J Med Virol* 2011;83:1301–1307.

12. Machado LF, Ishak MO, Vallinoto AC, Lemos JA, Azevedo VN *et al.* Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon Region of Brazil. *AIDS Res Hum Retroviruses* 2009;25:961–966.

13. Monteiro JP, Alcantara LC, de Oliveira T, Oliveira AM, Melo MA *et al.* Genetic variability of human immunodeficiency virus-1 in Bahia State, Northeast, Brazil: high diversity of HIV genotypes. *J Med Virol* 2009;81:391–399.

14. Guimarães ML, dos Santos Moreira A, Loureiro R, Galvão-Castro B, Morgado MG. High frequency of recombinant genomes in HIV type 1 samples from Brazilian southeastern and southern regions. *AIDS Res Hum Retroviruses* 2002;18:1261–1269.

15. Eyer-Silva WA, Morgado MG. Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. *Mem Inst Oswaldo Cruz* 2007;102:809–815.

16. Sabino EC, Shpaer EG, Morgado MG, Korber BT, Diaz RS *et al.* Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 1994;68:6340–6346.

17. Cavalcanti AM, Brito AM, Salustiano DM, Lima KO, Silva SP *et al.* Primary resistance of HIV to antiretrovirals among individuals recently diagnosed at voluntary counselling and testing centres in the metropolitan region of Recife, Pernambuco. *Mem Inst Oswaldo Cruz* 2012;107:450–457.

18. Guimarães ML, Couto-Fernandez JC, Eyer-Silva WA, Teixeira SL, Chequer-Fernandez SL *et al.* Analysis of HIV-1 BF *pr/rt* recombinant strains from Rio de Janeiro/Brazil reveals multiple unrelated mosaic structures. *Infect Genet Evol* 2010;10:1094–1100.

19. Véras NM, Véras VS, Ramalho ED, Kyaw C, Silva RR *et al.* HIV type 1 genetic variability in central Brazil. *AIDS Res Hum Retroviruses* 2007;23:1481–1490.

20. Vicente AC, Otsuki K, Silva NB, Castilho MC, Barros FS *et al.* The HIV epidemic in the Amazon Basin is driven by prototypic and recombinant HIV-1 subtypes B and F. *J Acquir Immune Defic Syndr* 2000;23:327–331.

21. Pessôa R, Watanabe JT, Calabria P, Alencar CS, Loureiro P *et al.* Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. *Transfusion* 2015;55:980–990.

22. Brasil Ministério da Saúde. *Boletim Epidemiológico HIV Aids; Year I, No. 1.* Brasilia: Brasil Ministério da Saúde; 2012. www.aids.gov.br/pagina/boletim-epidemiologico.

23. Cavalcanti AM, Lacerda HR, Brito AM, Pereira S, Medeiros D *et al.* Antiretroviral resistance in individuals presenting therapeutic failure and subtypes of the human immunodeficiency virus type 1 in the Northeast region of Brazil. *Mem Inst Oswaldo Cruz* 2007;102:785–792.

24. Ferreira V, Moura P, Crovella S, Sobhie Diaz R, Castelo Filho A *et al.* The influence of HIV-1 subtype in the response to therapeutic dendritic cell vaccine. *Open AIDS J* 2012;6:289–292.

25. de Medeiros LB, Lacerda HR, Cavalcanti AM, de Albuquerque MF. Primary resistance of human immunodeficiency virus type 1 in a reference center in Recife, Pernambuco, Brazil. *Mem Inst Oswaldo Cruz* 2006;101:845–849.

26. Santos LA, Monteiro-Cunha JP, Araujo AF, Brites C, Galvao-Castro B *et al.* Detection of distinct human immunodeficiency virus type 1 circulating recombinant forms in Northeast Brazil. *J Med Virol* 2011;83:2066–2072.

27. Carr JK, Avila M, Gomez Carrillo M, Salomon H, Hierholzer J *et al.* Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *Aids* 2001;15:F41–F47.

28. Bello G, Aulicino PC, Ruchansky D, Guimarães ML, Lopez-Galindez C *et al.* Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. *Retrovirology* 2010;7:22.

29. Ramos A, Tanuri A, Schechter M, Rayfield MA, Hu DJ *et al.* Dual and recombinant infections: an integral part of the HIV-1 epidemic in Brazil. *Emerg Infect Dis* 1999;5:65–74.

30. Aulicino PC, Holmes EC, Rocco C, Mangano A, Sen L. Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina. *J Virol* 2007;81:427–429.

31. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol* 2015;30:296–307.

32. Arenas M, Lorenzo-Redondo R, Lopez-Galindez C. Influence of mutation and recombination on HIV-1 *in vitro* fitness recovery. *Mol Phylogenet Evol* 2016;94:264–270.

33. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882.

34. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999;41:95–98.

35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.

36. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–267.

37. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999;73:152–160.

38. Martin DP, Williamson C, Posada D. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 2005;21:260–262.

39. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 2007;176:1035–1047.

40. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol* 1992;34:126–129.

41. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 2001;98:13757–13762.

42. Sawyer S. Statistical tests for detecting gene conversion. *Mol Biol Evol* 1989;6:526–538.

43. Bay RA, Bielawski JP. Recombination detection under evolutionary scenarios relevant to functional divergence. *J Mol Evol* 2011;73:273–286.

44. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* 2004;304:581–584.

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

46. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;29:1969–1973.

47. Yang Z, Rannala B. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst Biol* 2005;54:455–470.

48. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005;22:1208–1222.

49. Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 1994; 11:715–724.

50. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005;21:676–679.

51. Hué S, Clewley JP, Cane PA, Pillay D. HIV-1 *pol* gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *Aids* 2004;18:719–728.

52. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE *et al.* The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA* 2007;104:18566–18570.

53. Junqueira DM, de Medeiros RM, Matte MC, Araújo LA, Chies JA *et al.* Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One* 2011;6:e27489.

54. Arruda E, Simões L, Sucupira C, Medeiros M, Arruda E *et al.* Short communication: intermediate prevalence of HIV type 1 primary antiretroviral resistance in Ceará state, Northeast Brazil. *AIDS Res Hum Retroviruses* 2011;27:153–156.

55. Couto-Fernandez JC, Morgado MG, Bongertz V, Tanuri A, Andrade T *et al.* HIV-1 subtyping in Salvador, Bahia, Brazil: a city with African sociodemographic characteristics. *J Acquir Immune Defic Syndr* 1999;22:288–293.

56. Gadelha SR, Shindo N, Cruz JN, Morgado MG, Galvão-Castro B. Molecular epidemiology of human immunodeficiency virus-1 in the State of Ceará, Northeast, Brazil. *Mem Inst Oswaldo Cruz* 2003; 98:461–464.

57. Pedroso C, Queiroz AT, Alcântara LC, Drexler JF, Diaz RS *et al.* High prevalence of primary antiretroviral resistance among HIV-1-infected adults and children in Bahia, a Northeast State of Brazil. *J Acquir Immune Defic Syndr* 2007;45:251–253.

58. Alencar CS, Sabino EC, Carvalho SM, Leao SC, Carneiro-Proietti AB *et al.* HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: role of infected blood donors as sentinel populations for molecular surveillance of HIV. *J Acquir Immune Defic Syndr* 2013;63:387–392.

59. Aulicino PC, Bello G, Rocco C, Romero H, Mangano A *et al.* Description of the first full-length HIV type 1 subtype F1 strain in Argentina: implications for the origin and dispersion of this subtype in South America. *AIDS Res Hum Retroviruses* 2007; 23:1176–1182.

60. Thomson MM, Sierra M, Tanuri A, May S, Casado G *et al.* Analysis of near Full-Length genome sequences of HIV type 1 BF intersubtype recombinant viruses from Brazil reveals their independent origins and their lack of relationship to CRF12_BF. *AIDS Res Hum Retroviruses* 2004;20:1126–1133.

61. Aulicino PC, Bello G, Guimaraes ML, Ruchansky D, Rocco C *et al.* Longitudinal analysis of HIV-1 BF1 recombinant strains in vertically infected children from Argentina reveals a decrease in CRF12_BF *pol* gene mosaic patterns and high diversity of BF unique recombinant forms. *Infect Genet Evol* 2011;11:349–357.

62. Brennan CA, Brites C, Bodelle P, Golden A, Hackett J Jr *et al.* HIV-1 strains identified in Brazilian blood donors: significant prevalence of B/F1 recombinants. *AIDS Res Hum Retroviruses* 2007;23: 1434–1441.

63. Brígido LF, Nunes CC, Oliveira CM, Knoll RK, Ferreira JL *et al.* HIV type 1 subtype C and CB *pol* recombinants prevail at the cities with the highest AIDS prevalence rate in Brazil. *AIDS Res Hum Retroviruses* 2007;23:1579–1586.

64. Gräf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 2013;435:170–178.

65. Alcalde R, Guimarães ML, Duarte AJ, Casseb J. Clinical, epidemiological and molecular features of the HIV-1 subtype C and recombinant forms that are circulating in the city of São Paulo, Brazil. *Virol J* 2012;9:156.

66. Brígido LF, Ferreira JL, Almeida VC, Rocha SQ, Ragazzo TG *et al.* Southern Brazil HIV type I C expansion into the state of São Paulo, Brazil. *AIDS Res Hum Retroviruses* 2011;27:339–344.

67. Carvalho BC, Cardoso LP, Damasceno S, Stefani MM. Moderate prevalence of transmitted drug resistance and interiorization of HIV type 1 subtype C in the inland North state of Tocantins, Brazil. *AIDS Res Hum Retroviruses* 2011;27:1081–1087.

68. Delatorre EO, Bello G, Eyer-Silva WA, Chequer-Fernandez SL, Morgado MG *et al.* Evidence of multiple introductions and autochthonous transmission of the HIV type 1 CRF02_AG clade in Brazil. *AIDS Res Hum Retroviruses* 2012;28:1369–1372.

69. Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH *et al.* HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. *PLoS One* 2014;9:e87622.

70. Bello G, Guimarães ML, Morgado MG. Evolutionary history of HIV-1 subtype B and F infections in Brazil. *Aids* 2006;20:763–768.

71. Abecasis AB, Wensing AM, Paraskevis D, Vercauteren J, Theys K *et al.* HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* 2013;10:7.

72. Hawke KG, Waddell RG, Gordon DL, Ratcliff RM, Ward PR *et al.* HIV non-B subtype distribution: emerging trends and risk factors for imported and local infections newly diagnosed in South Australia. *AIDS Res Hum Retroviruses* 2013;29:311–317.

73. Li L, Chen L, Yang S, Li T, Li J *et al.* Recombination form and epidemiology of HIV-1 unique recombinant strains identified in Yunnan, China. *PLoS One* 2012;7:e46777.

74. Neogi U, Sood V, Banerjee S, Ghosh N, Verma S *et al.* Global HIV-1 molecular epidemiology with special reference to genetic analysis of HIV-1 subtypes circulating in North India: functional and pathogenic implications of genetic variation. *Indian J Exp Biol* 2009;47:424–431.

75. Soares EA, Martínez AM, Souza TM, Santos AF, da Hora V *et al.* HIV-1 subtype C dissemination in Southern Brazil. *Aids* 2005;19: S81–S86.

76. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL *et al.* Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol* 2007;7:263–270.

77. Leal ES, Holmes EC, Zanotto PM. Distinct patterns of natural selection in the reverse transcriptase gene of HIV-1 in the presence and absence of antiretroviral therapy. *Virology* 2004;325: 181–191.

78. Arenas M. Genetic consequences of antiviral therapy on HIV-1. *Comput Math Methods Med* 2015;2015:395826:1–9.