



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

Rodrigo Pessôa¹, Jaqueline Tomoko Watanabe¹, Paula Calabria¹, Alvina Clara Felix¹, Paula Loureiro², Ester C. Sabino³, Michael P. Busch⁴, Sabri S. Sanabani^{5*} for the International Component of the NHLBI Recipient Epidemiology and Donor Evaluation Study-III (REDS-III)¹

1 Virology Department, São Paulo Institute of Tropical Medicine, University of São Paulo, São Paulo, Brazil, **2** Pernambuco State Center of Hematology and Hemotherapy, Recife, Pernambuco, Brazil, **3** Department of Infectious Disease/Institute of Tropical Medicine, University of São Paulo, São Paulo, Brazil, **4** Blood Systems Research Institute, San Francisco, California, United States of America, **5** Clinical Laboratory, Department of Pathology, Hospital das Clínicas, School of Medicine, University of São Paulo, São Paulo, Brazil

Abstract

Background: The findings of frequent circulation of HIV-1 subclade F1 viruses and the scarcity of BF1 recombinant viruses based on *pol* subgenomic fragment sequencing among blood donors in Pernambuco (PE), Northeast of Brazil, were reported recently. Here, we aimed to determine whether the classification of these strains (n = 26) extends to the whole genome sequences.

Methods: Five overlapping amplicons spanning the HIV near full-length genomes (NFLGs) were PCR amplified from peripheral blood mononuclear cells (PBMCs) of 26 blood donors. The amplicons were molecularly bar-coded, pooled, and sequenced by Illumina paired-end protocol. The prevalence of viral variants containing drug resistant mutations (DRMs) was compared between plasma and PBMCs.

Results: Of the 26 samples studied, 20 NFLGs and 4 partial fragments were *de novo* assembled into contiguous sequences and successfully subtyped. Two distinct BF1 recombinant profiles designated CRF70_BF1 and CRF71_BF1, with 4 samples in profile I and 11 in profile II were detected and thus constitute two novel recombinant forms circulating in PE. Evidence of dual infections was detected in four patients co-infected with distinct HIV-1 subtypes. According to our estimate, the new CRF71_BF1 accounts for 10% of the HIV-1 circulating strains among blood donors in PE. Discordant data between the plasma and PBMCs-virus were found in 15 of 24 donors. Six of these strains displayed major DRMs only in PBMCs and four of which had detectable DRMs changes at prevalence between 1-20% of the sequenced population.

Conclusions: The high percentage of the new RF71_BF1 and other BF1 recombinants found among blood donors in Pernambuco, coupled with high rates of transmitted DRMs and dual infections confirm the need for effective surveillance to monitor the prevalence and distribution of HIV variants in a variety of settings in Brazil.

Citation: Pessôa R, Watanabe JT, Calabria P, Felix AC, Loureiro P, et al. (2014) 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 9(11): e112674. doi:10.1371/journal.pone.0112674

Editor: Lars Kaderali, Technische Universität Dresden, Medical Faculty, Germany

Received: May 20, 2014; **Accepted:** October 10, 2014; **Published:** November 17, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All consensus genome assemblies generated in this study were submitted to NCBI's GenBank database (KJ849757-KJ849783).

Funding: This work was supported by grants 2011/11090-5 and 2011/12297-2 from the Fundação de Amparo à Pesquisa do Estado de São Paulo. The study was also supported by the Retrovirus Epidemiology Donor Study-II (HHSN268200417175C) and the Recipient Epidemiology and Donor Outcomes Study-III (HHSN268201100007I) contracts from NHLBI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* Email: sabyem_63@yahoo.com

¶ Those responsible for the Recipient Epidemiology and Donor Outcomes Study-III are listed in the acknowledgments.

Introduction

One of the most prominent features of HIV-1 is the remarkable accumulation of genetic diversity in its population during the course of infection. This diversity can be attributed to the high mutation rate of reverse transcriptase (3×10^{-5} substitutions per site per generation) [1], rapid viral turnover (10^8 to 10^9 virions per day) [2], large number of infected cells (10^7 to 10^8 cells) [3], and recombination events that are taking place during replication [4]. Consequently, the HIV-1 population is composed of a swarm of highly genetically related variants, i.e. a *quasispecies*, capable of rapidly adapting to various selective pressures. This diversity has been shown to have an impact not only on viral phenotypes at the level of transmission patterns, pathogenicity and immunology but also in responses to antiretroviral therapy and vaccines [5,6]. Nine distinct genetic subtypes, (A–D, F–H, J and K) are joined in the pandemic today by more than 70 major circulating recombinant forms (CRFs) [<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>] and numerous unique recombinant forms (URFs) have been isolated from individual patients [7]. Recombination between the URFs and CRFs and between the existing CRFs (inter-CRF recombinants) results in emergence of novel second and third generation recombinant forms which would further continue to shape the future of HIV epidemic through the generation of other variants with improved fitness to influence viral transmissibility [8,9]. It has been reported that recombinant viruses including the URFs and CRFs may account for at least 20% of all HIV infections [10]. The existence of recombinant viruses is an evidence of simultaneous infection of multiple viruses during a single transmission event (co-infection) or from the sequential infection of viruses during multiple transmission events (superinfection).

Brazil, the most populous country in the Latin America, is home to about one third of the people living with HIV (608,230 in Central and South America (UNAIDS. 2010–2011 Report on the Brazilian response to HIV/AIDS). HIV-1 subtype B is a major genetic clade circulating in the country but the overall prevalence of non-B strains, particularly URF BF1, C and URF BC, is increasing [11,12,13,14]. Data from recent studies of the viral near full-length genomes (NFLGs) have provided evidence of Brazilian CRF strains designated as CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF, CRF46_BF and CRF31_BC [12,13,15,16]. Recently, Alencar et al [17] performed a molecular epidemiological survey of HIV-1 in Brazil and analyzed the partial *pol* gene of 341 samples from seropositive blood donors collected between 2007 and 2011 from 4 Brazilian blood centers of the REDS-II (Retrovirus Epidemiological Donor Study) international program. The study reported a relatively high prevalence of subclade F1 (26 [24%] of 110), and only one case of BF1 recombinant among blood donors from Recife, Pernambuco (PE). These findings contrast with those from the previous studies of HIV-1 NFLGs in Brazil [13,18,19,20,21]. We therefore undertook this study to determine whether the classification of these strains extends to the whole genome sequences. Additionally, we aimed to compare the rate of drug resistance mutations (DRMs) in plasma bulk RNA and PBMC massively parallel sequencing (MPS) data to elucidate the differences in resistance profile between both compartments. The results revealed a considerable diversity of BF1 mosaic structures and high percentage of the new RF71_BF1 and other BF1 recombinants among blood donors in PE with high rates of transmitted DRMs and dual infections.

Materials and Methods

Study samples

The 24 peripheral blood mononucleated cells (PBMCs) samples reported in this study were from HIV-1 seropositive blood donors in Recife, capital of Pernambuco in the northeast region of Brazil. The rationale for collection of these samples has been previously reported [17]. No evidence of direct epidemiological linkage could be established. All study subjects provided written informed consent. Ethical approval was obtained from the local ethical review committee of the HEMOPE foundation as well as the Recipient Epidemiology and Donor Evaluation Study-II collaborating centers (Blood Systems Research Institute/University of California San Francisco) and Data Coordinating Center (Westat, Inc.) in the US.

DNA extraction and amplification of the NFLGs

The genomic DNA used for the PCR analyses was extracted using the QIAamp blood kit (Qiagen GmbH, Hilden Germany) according to the manufacturer's instructions. The NFLGs from five overlapping fragments were obtained by PCR using the Platinum *Taq* DNA Polymerase High Fidelity (5 U/ μ l) (Invitrogen, Life Technologies, Carlsbad, CA) and determined by a previously reported method [12,18]. The amplified DNA fragments from the nested PCR products were separated by gel electrophoresis and purified using Freeze 'N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad, Hercules, CA, USA). Each purified amplicon was quantified using Quant-IT HS reagents (Invitrogen, Life Technologies, Carlsbad, CA), and all five amplicons from a single viral genome were pooled together at equimolar ratios.

Whole viral genome library preparation

Sequencing libraries were prepared as described previously [22]. Briefly, one ng of each sample amplicon pool was used in a fragmentation reaction mix using the Nextera XT DNA sample prep kit according to the manufacturer's protocol (Illumina, San Diego, CA). The tagmentation and fragmentation of each pool were simultaneously performed by incubation for 5 min at 55°C followed by incubation in neutralizing tagment buffer for 5 min at room temperature. After neutralization of the fragmented DNA, a light 12-cycle PCR was performed with Illumina Ready Mix to add Illumina flowcell adaptors, indexes and common adapters for subsequent cluster generation and sequencing. Amplified DNA library was then purified using Agencourt AMPure XP beads (Beckman Coulter), which excluded very short library fragments. Following AMPure purification, the quantity of each library was normalized to ensure equally library representation in our pooled samples. Prior to cluster generation, normalized libraries were further quantified by realtime PCR (qPCR) using the SYBR fast Illumina library quantification kit (KAPA Biosystems) following the instructions of the manufacturer. The qPCR was run on the 7500 Fast Real-Time PCR System (Applied Biosystems). The thermocycling conditions consisted of an initial denaturation step at 95°C for 5 min followed by 35 cycles of 30 s at 95°C and 45 s at 60°C. The final libraries were pooled at equimolar concentration and diluted to 4 nM. To denature the indexed DNA, 5 μ L of the 4 nM library were mixed with 5 μ L of 0.2 N fresh NaOH and incubated for 5 min at room temperature. 990 μ L of chilled Illumina HT1 buffer was added to the denatured DNA and mixed to make a 20 pM library. After this step, 360 μ L of the 20 pM library was multiplexed with 6 μ L of 12.5 pM denatured PhiX control to increase sequence diversity and then mixed with 234 μ L of chilled HT1 buffer to make a 12 pM sequenceable library.

Finally, 600 μ L of the prepared library was loaded on an Illumina MiSeq clamshell style cartridge for paired end 250 bp sequencing reads.

Data analysis

Fastq files were generated by the Illumina MiSeq reporter for downstream analysis and validated to evaluate the distribution of quality scores and to ensure that quality scores do not drastically drop over each read. To take the sequencing error rate into account, we only considered variants detected at a frequency higher than 1% and Phred quality score of >30%, i.e., a base call accuracy of 99.9%. Validated fastq files from each viral genome were *de novo* assembled into contiguous sequences and annotated with CLC Genomics Workbench version 5.5 (CLC Bio, Aarhus, Denmark) with default parameters and were additionally assembled using Velvet implemented in the Sequencher program 5.2 (Gene Code Corp., Ann Arbor, MI). The contiguous genomic sequence from each virus strain was extracted from the assembly and used for further analysis. The full designation of samples, according to WHO-proposed nomenclature, is YYBR_PEXXX, where YY stands for the year of study, BR for Brazil, PE for Pernambuco, and XXX stands for sample number.

Screening for recombination events and identification of breakpoints

The *de novo* assembled NFLGs and partial consensus sequences were aligned with reference sequences representing subtypes A–D, F–H, J and K obtained from the Los Alamos database (<http://hiv-web.lanl.gov>) using MAFFT version 7 [23]. Aligned sequences were manually edited and trimmed to the minimal shared length in the BioEdit Sequence Alignment Editor Program. The gap-stripped aligned sequences were screened for the presence of recombination by the bootscan methods and similarity plots implemented in the SIMPLOT program v3.5.1 [24,25]. The following parameters were used in bootscan method: window size, 350 bp; step size, 30 bp; the F84 model of evolution (Maximum likelihood (ML)) as a model to estimate nucleotide substitution; transition/transversion ratio, 2.0; and a bootstrap of 100 trees. In addition, the significant threshold for the bootscan was set at 70%. The jumping profile Hidden Markov Model (jpHMM) [26] was also used to confirm the recombination events and to define the recombination breakpoints according to the HXB2 coordinate system. Recombinant regions of the alignment as determined by the crossover points from the jpHMM and bootscan were analyzed separately by phylogenetic analysis. In further analysis, a network reconstruction was performed for the data set with evidence of recombination using SplitsTree4 version 4.3 [27] using the Neighbor-Net method. The NeighborNet method and the GTR+I+G distance model were used to create the network.

Phylogenetic Analysis

ML phylogenies were constructed using the GTR+I+G substitution model and a BIONJ starting tree. Heuristic tree searches under the ML optimality criterion were performed using the NNI branch-swapping algorithm. The approximate likelihood ratio test (aLRT) based on a Shimodaira-Hasegawa-like procedure was used as a statistical test to calculate branch support. The maximum composite likelihood in MEGA version 6 [28] was used to calculate the genetic distances between and within isolates. All trees were displayed using MEGA version 6.0 software.

Genotyping Analysis

For provirus sequences generated in this study, the MPS reads of partial *pol* gene associated with DRMs in the protease and reverse transcriptase regions of the HIV-1 genome of each sample were aligned to their corresponding consensus sequence using the CLC Genomics Workbench version 5.5 (CLC Bio, Aarhus, Denmark). The minority HIV-1 resistant variants were identified using a threshold of >1.0% of the reads sequenced. Reads with < 1% were discarded to account for potential errors due to the error rate of PCR. Amino acid positions including all listed major mutations and minor mutations associated with drug resistance were identified according to the IAS-USA 2011 and Stanford HIV drug resistance database. Due to the polymorphic nature of most minor protease substitutions, we only considered major mutations as evidence of transmission of drug resistance. The list of the NRTIs resistance related sites **41**, 62, **65**, **67**, **69**, **70**, 71, 74, 75, 77, 115, 116, **151**, **184**, **210**, **215** and **219**; NNRTIs resistance related sites 90, 98, **100**, **101**, **103**, **106**, 108, 138, 179, **181**, **188**, **190**, 221, 225, 227, **230**; PIs resistance related sites 10, 16, 11, 20, 24, **30**, **32**, **33**, 35, 36, 43, **46**, **47**, **48**, **50**, 53, **54**, **58**, 60, 62, 63, 64, 69, 71, 73, **74**, **76**, 77, **82**, 83, **84**, 85, **88**, 89, **90** and 93. Bolded numbers are major drug resistance mutation sites.

GenBank accession numbers

All consensus genome assemblies generated in this study were submitted to NCBF's GenBank database (KJ849757 - KJ849783). The two distinct profiles of BF1 identical recombinant structure were registered with the Los Alamos National Database as CRF70_BF1 and CRF71_BF1.

Results

A total of 26 strains preliminarily classified as subclade F1 (n = 25) and BF1 recombinant (n = 1) by sequence analysis of a partial *pol* region in a previous study [17] were corroborated by further phylogenetic analysis of the NFLGs and larger partial fragments. Of note, one BF1 sample (10BR_PE059) was erroneously classified as F1 in our previous study. This sequence was thus corrected in this study and used for further analyses. Sequences were obtained for all five overlapped fragments that cover the NFLGs of 20 participants. Partial sequences were obtained from four fragments derived from 4 samples as shown in **Table 1**. Two samples did not amplify for any fragment. This might be a result of technical difficulties in recovering the provirus, but it is also possible that cells other than PBMCs are infected during the early stages of HIV infection. For the purposes of this report, these 2 samples were not considered further, and the analyses include only the 24 samples whose subtypes were successfully determined.

Analysis of the proviral NFLGs and partial consensus sequences revealed all isolates retain intact reading frames for a majority of their genes and no gross deletions or rearrangements were observed. Recombination analysis from each strain shows two distinct recombinant profiles with 4 samples in profile I and 11 in profile II. The recombinant genomes of both profiles essentially consisted of subclades F1 and subtype B as parental sequences and appeared different from all previously documented CRFs in Brazil and South America. Plausible breakpoints identified in each profile using bootscanning analyses were consistent with those identified using jpHMM (schematically illustrated in Figure 1 and 2). Therefore, the new recombinant strains in profile I and II are now designated CRF70_BF1 and CRF71_BF1, respectively. All of the breakpoints in the two CRFs were mapped and compared. The results revealed great similarity in structure between the two

Table 1. The near full-length genomic (NFLG) and partial fragments subtyping of HIV-1 from plasma and blood samples.

Sample ID	Sequence Fragment				Subtype		Number of Reads	Av. coverage	
	A ₍₅₄₆₋₂₅₉₈₎	B1 ₍₂₁₅₇₋₃₇₉₁₎	B2 ₍₃₂₃₆₋₅₂₂₀₎	C ₍₄₈₉₀₋₇₈₀₈₎	D ₍₇₇₁₉₋₉₅₃₇₎	Plasma			Provirus
10BR_PE002	-	+	+	+	+	F	BF1	145757	21,299
10BR_PE004	+	+	+	+	+	F	BF1_CRF70	30843	2625
10BR_PE008	+	+	+	+	+	F	BF1_CRF71	111362	2564
10BR_PE009	+	+	+	+	+	F	BF1_CRF71	23923	5624
10BR_PE016	+	+	+	+	+	F	BF1_CRF70	29409	584
10BR_PE025	+	+	+	+	+	F	BF1_CRF70	84718	2173
10BR_PE026	+	+	-	+	+	F	BF1_CRF71	20503	2635
10BR_PE032	+	+	+	-	+	F	BF1	55741	6084
10BR_PE053	+	+	+	+	+	F	B	78936	3697
10BR_PE059 ¹	+	+	+	+	+	F	BF1	30166	2619
10BR_PE064	+	+	+	+	+	F	BF1_CRF71	219755	5608
10BR_PE066	+	+	+	+	+	F	BF1_CRF71	32024	6701
10BR_PE071	+	+	+	+	+	F	BF1_CRF71	98156	5578
10BR_PE073	+	+	+	+	-	F	BF1	47041	1365
10BR_PE084	+	+	+	+	+	F	BF1_CRF71	15195	2034
10BR_PE085	-	-	-	-	-	F	-	-	-
10BR_PE086	+	+	+	+	+	F	BF1	15348	1106
10BR_PE087	-	-	-	-	-	F	-	-	-
10BR_PE088	+	+	+	+	+	F	BF1_CRF71	42801	700
10BR_PE090	+	+	+	+	+	F	BF1_CRF71	18052	187
10BR_PE092	+	+	+	+	+	F	BF1_CRF71	4546	201
10BR_PE094	+	+	+	+	+	F	BF1_CRF71	273899	6870
10BR_PE102	+	+	+	+	+	F	BF1	494376	7257
10BR_PE104	+	+	+	+	+	BF1	B/BF1	26901	3324
10BR_PE107	+	+	+	+	+	F	F1	78655	2159
10BR_PE109	+	+	+	+	+	F	BF1_CRF70	165634	6190

¹Variant was erroneously classified as F1 in our previous study. doi:10.1371/journal.pone.0112674.t001

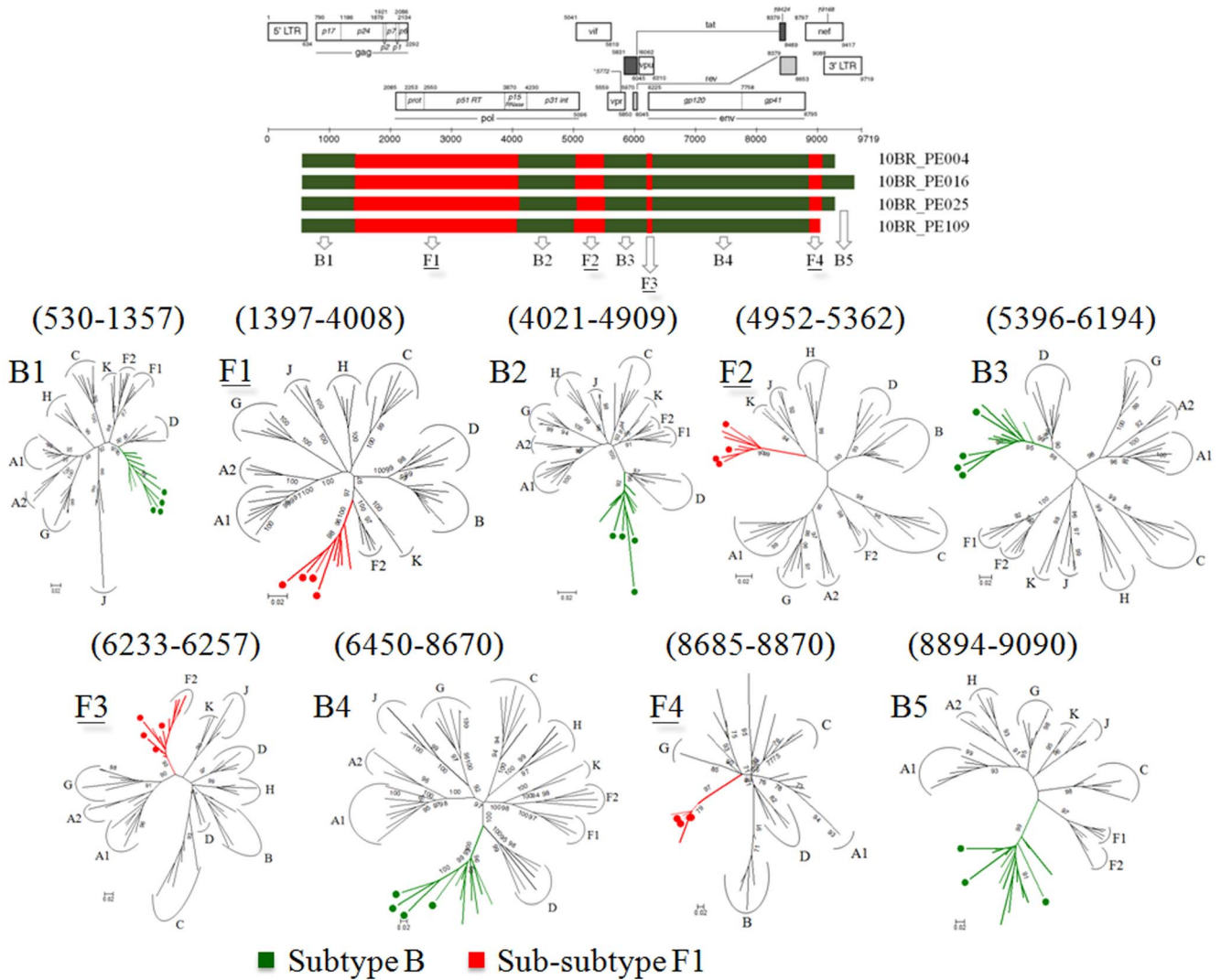


Figure 1. Schematic representation of NFLGs structure and breakpoint profiles with confirmatory phylogenetic trees of the four sequences identified in this study as CRF70_BF1 (colored circles). The phylogenetic trees of the nine mosaic segments defined by jpHMM, similarity plot and bootscan analysis were constructed with PHYML v.3.0. For clarity purposes, the trees were midpoint rooted. The region of subclade F1 and subtypes B are as indicated at the bottom. Positions of breakpoints are numbered relative to the HXB2 numbering system. The approximate likelihood ratio test (aLRT) values of $\geq 90\%$ are indicated at nodes. The scale bar represents 0.02 nucleotide substitution per site. The results from the ML analysis were sufficiently robust to confirm the structure for the four specimens that were suggested by the recombination analysis. doi:10.1371/journal.pone.0112674.g001

CRFs. Among the sub-subtype F1 stretches, identified by letter F1 through F4 and F1 through F3 in CRF70_BF1 and CRF71_BF1, respectively (Figure 1 and 2), fragments F1, F2 and F4 appear to be in the same locations. For patient 10BR_PE026, we were unable to amplify fragment designated as B2 in **Table 1** (position 3236–5220) and therefore could not investigate the location of breakpoint at the 5' of fragment B2 as depicted in Figure 2. Different from CRF71_BF1, CRF70_BF1 appears to contain one too short sub-subtype F1 fragment within the *Vpu* gene (34 bp, nucleotide 6233–6266 according to position in HXB2 GenBank accession no. K03455) interspersed with subtype B. In attempt to find sequences with recombinant structure similar to any of the strains in the two CRFs, multiple sequence alignments including all published BF1 unique NFLGs sequences were built, aligned and subjected to bootscan and automated jpHMM. The results showed that the CRF71_BF1 sequences had a mosaic sequence pattern nearly identical to the previously published Brazilian BF1

isolate 02BR033 (GenBank: DQ358811) in sample collected in 2002 from patient living in São Paulo, Southeast of Brazil [13]. To further test for recombination, ML phylogenetic trees were inferred for the regions of nucleotide sequence on either side of the breakpoints detected by bootscan and jpHMM methods. As shown in Figure 1 and 2, the exploratory tree analysis revealed that each fragment of recombinant virus from each CRF clustered tightly ($>90\%$ aLRT) with corresponding fragments of subtype B or F1 reference sequences in agreement with the subtype assigned by recombination analysis. The only disagreement between the recombination and exploratory tree analyses was for a short CRF70_BF1 fragment in the middle of the *Vpu* region (F3 in Figure 2). Despite the fact that F3 fragment in CRF70_BF1 strains have shorter sequences and some group M variants cannot resolve some of the internal nodes, all of the CRF70_BF1 can resolve the terminal nodes and appeared more closely related phylogenetically to subclade F1 than the other groups of reference sequences. To

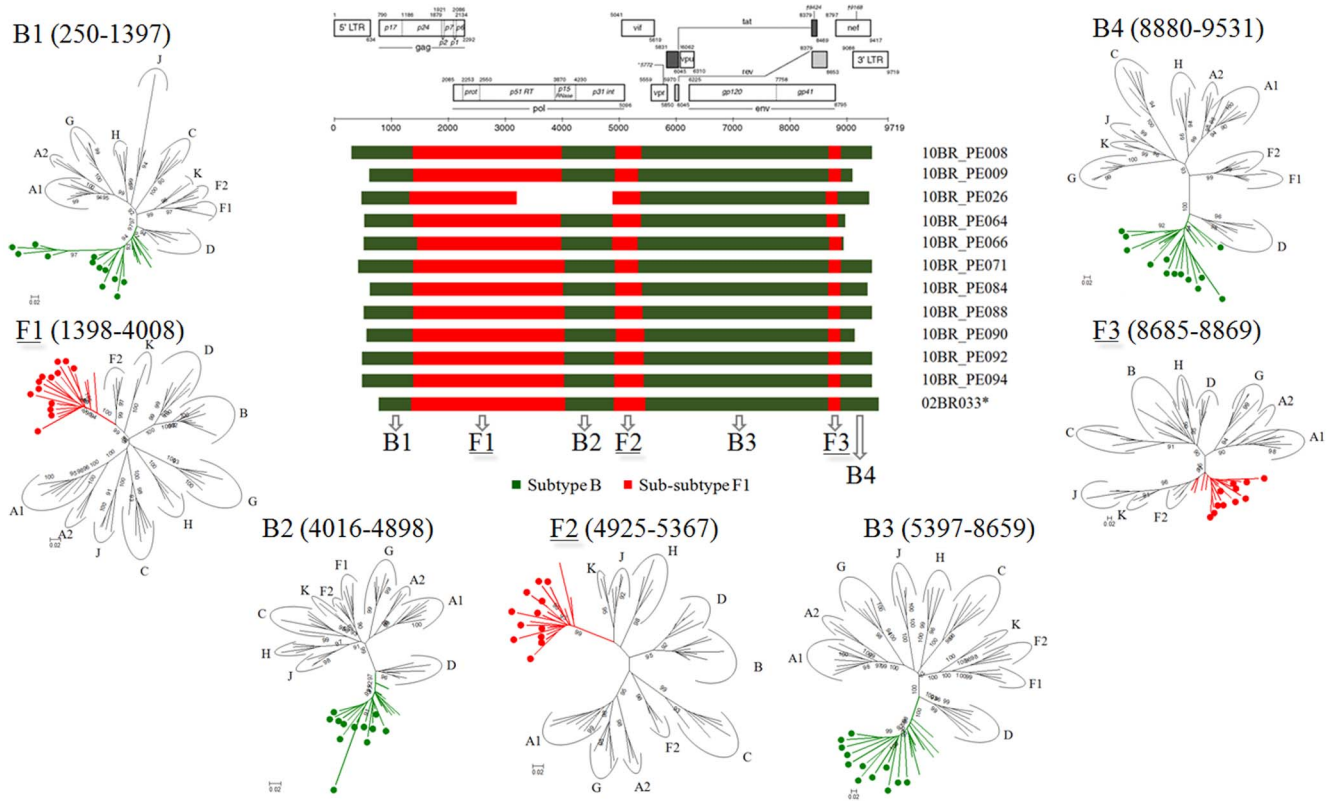


Figure 2. Schematic representation of NFLGs structure and breakpoint profiles with confirmatory phylogenetic trees of the twelve sequences identified in this study as CRF70_BF1 (colored circles). For patient 10BR_PE026, we were unable to amplify fragment designated as B2 and therefore we could not investigate the location of breakpoint at the 5' of fragment B2. All ML phylogenetic trees were constructed using the PHYML v.3.0 package. The region of subclade F1 and subtypes B are indicated at the bottom. Positions of breakpoints are numbered relative to the HXB2 numbering system. For clarity purposes, the trees were midpoint rooted. The approximate likelihood ratio test (aLRT) values of $\geq 90\%$ are indicated at nodes. The scale bar represents 0.02 nucleotide substitution per site. The results from the ML analysis were sufficiently robust to confirm the structure for the twelve specimens that were suggested by the recombination analysis. doi:10.1371/journal.pone.0112674.g002

increase the phylogenetic signal, ML trees were performed on concatenated data of discontinuous subtype B and F1 fragment from CRF70_BF1 (Figure 3A and 3B) and CRF71_BF1 (Figure 3A and 3B), which was well confirmed by aLRT values above 90% as shown in Figure 3. Split decomposition was then used to visualize the relationship of the two CRFs (Figure 4). While sequences within each of these two CRF did not group directly with each other, they were present as one group of sequences within a cluster between subtype B and F1. The observed mean intrasubtype genetic distances were 8.2% (range 6.7–10.0%) and 8.7% (range 6.2–9.9%) for CRF70_BF1 and CRF71_BF1 strains, respectively, and closer to the intersubtype distance between both CRFs (8.3%).

Beside the CRFs, in five of the seven strains initially classified as infected with subclade F1, subtype B fragments were detected and appeared to represent URFs strains (Figure 5). The relationships of the viral sequences from patients' PBMCs to the sequences obtained from the corresponding RNA viruses within the same regions were examined for each patient to roughly assess the viral diversity in both compartments (Figure 6). Surprisingly, the intra-individual plasma and proviral sequence variation in four patients (10BR_PE073, 10BR_PE053, 10BR_PE104 and 10BR_PE032) in the partial *pol* regions depicted in Figure 5 (marked with orange boxes) were remarkably high, indicating that the plasma viruses were derived from a population significantly distinct from those of the cellular sources; a result consistent with dual infection with

different subtypes (Figure 5 and 6). Except those with dual infections, all the other sequences from both compartments were located close to one another on the same branch and had plasma RNA and proviral DNA variation only ranging between 0.1–0.8% (Figure 4). In the case of subject 10BR_PE104, MPS data revealed a mixture of two distinct consensus sequences, one NFLGs from strain B and a second recombinant of strains B and F1 (4450 bp) almost identical to the plasma virus in the same region. The BF1 segment was present in 7% of MPS reads with a median coverage of 250 reads in the sample from subject 10BR_PE104, while the NFLGs present in 93% of MPS reads with a median coverage of 7250 reads. Thus, the low BF1 copies would not have been detectable using the provirus traditional bulk sequencing approaches. Again, these data confirms that infection in this subject was founded by two genetic lineages. In Figure 5, there were three samples (10BR_PE002 (pro), 10BR_PE032 [17], and 10BR_PE102 [17], which shared one breakpoint at the integrase gene (position 2393–2462 nt), similar to the third breakpoint of the CRF71_BF variants (5' of fragment F2, Figure 2). This increased prevalence of genetic breakpoints in the integrase region may indicate a possible preference region for recombination to occur.

Three viruses that were classified as BF1 (n = 2; 10BR_PE059, 10BR_PE086) and F1 (10BR_PE107) on partial *pol* analysis were confirmed by NFLGs analysis (Figure 5).

As stated before, the 26 samples selected to this study are derived from 110 blood donors from Recife PE, and all were

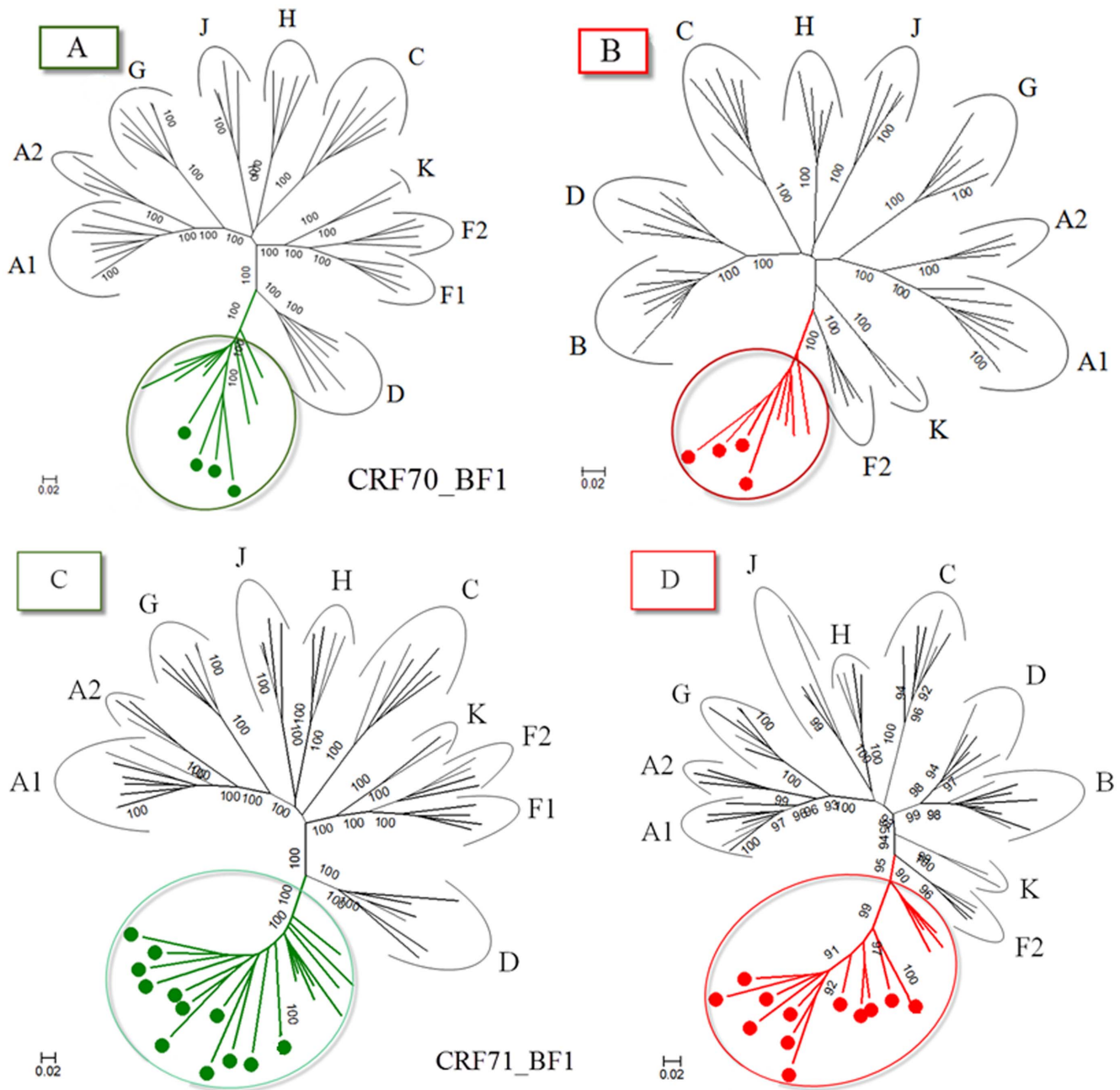


Figure 3. ML phylogenetic tree from concatenated regions assigned as subtype B and F1 from four CRF70_BF1 (3A and 3B) and twelve CRF71_BF1 (3C and 3D) isolates as defined by recombination analysis. Sequences from each region were aligned with reference sequences representing subtypes A–D, F–H, J and K obtained from the Los Alamos database (<http://hiv-web.lanl.gov>). For clarity purposes, the trees were midpoint rooted. The approximate likelihood ratio test (aLRT) values of $\geq 90\%$ are indicated at nodes. The scale bar represents 0.02 nucleotide substitution per site. The results from this analysis revealed that each segment of the CRF70_BF1 (3A and 3B) and twelve CRF71_BF1 (3C and 3D) viruses was found to cluster with corresponding segments of subtype B or F1 viruses in agreement with the subtype assigned by recombination analysis.

doi:10.1371/journal.pone.0112674.g003

identified as infected with subclade F1 except two donors who were found to be infected with BF1 recombinant viruses. Genotyping based on NFLGs or long partial fragments of these samples was successful in 24 donors, of whom 11 were infected with CRF71_BF1, 4 with CRF70_BF1, 6 BF1 URF, 1 subtype B, 1 subtype F1, and 1 dually infected with BF1 and B viruses. Based on this analysis, it is evident that non-recombinant subclade F1 accounts for <1% of HIV-1 subtype's circulating in PE and that

CRF71_BF1 is responsible for 50% (11/22) of infections caused by BF1 recombinants.

We also aimed in this study to analyze the emergence of proviral HIV DRMs detected by MPS and compare the results with the plasma HIV DRMs detected by previously described plasma bulk sequencing of the same blood donors group to better understand the resistance profile between the two compartments. A summary of these results are provided in **Table 2**. Discordant data between

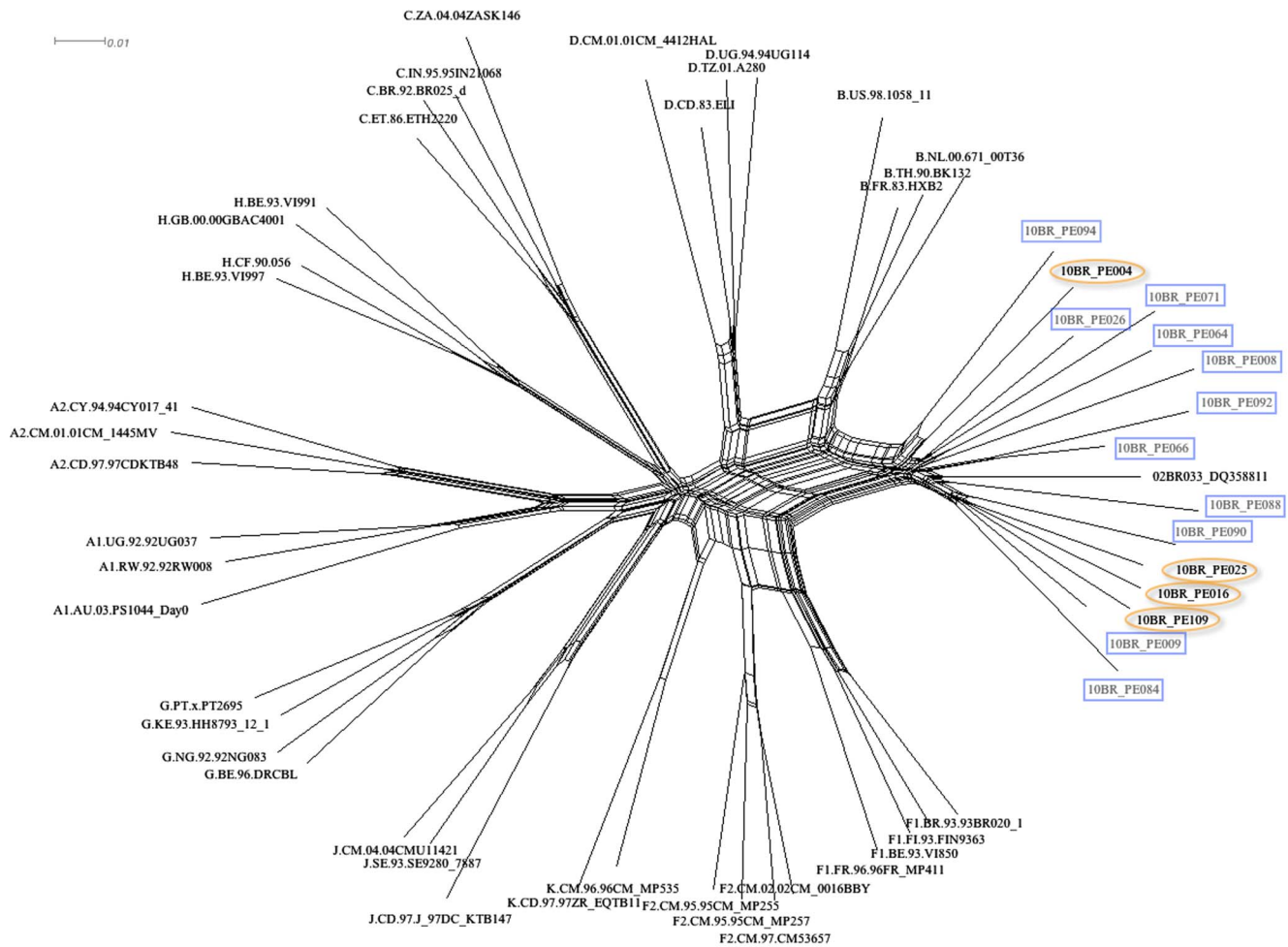


Figure 4. Split networks for four sequences of CRF_70BF1 (marked with orange circles) and eleven CRF_71BF1 (marked with blue boxes) and representatives of HIV subtypes A–K sequences from the Los Alamos database (<http://hiv-web.lanl.gov>). The splits graphs used distances computed under the GTR+I+G model. The scale bar represents 0.01 nucleotide substitution per site. The results demonstrate that sequences within each of CRF did not group directly with each other, but they were present as one group of sequences within a cluster between subtype B and subclade F1.
doi:10.1371/journal.pone.0112674.g004

the cell free viruses and the PBMC-viruses were found in 15 participants. Six of these strains displayed major amino-acid changes only in the cellular compartment and four of which had detectable major amino-acid changes at prevalence between 1–20% of the sequenced population. Neither major amino-acid changes in the protease nor in the reverse transcriptase coding region were detectable by plasma bulk sequencing. Eleven and two DRMs in the protease and reverse transcriptase regions, respectively, were detected in plasma bulk sequencing but were not observed in MPS data. For the coding region of the protease gene, major PI resistance mutations, namely, M46I, and L33F were detected in four blood donors. As shown in **Table 2**, the overall frequency of minor mutations in both compartments in the protease gene was significantly greater than that detected in reverse transcriptase. Although these mutations were mostly polymorphisms and not directly responsible for drug resistance but they are able to compensate the low protease processivity caused by primary mutations [29]. The proviral DNA MPS analysis in the reverse transcriptase region showed major mutations at the following codons in five patients: M184I, M230I, and E138K. It is well known that HIV-1 harboring the M184I/V mutations has a low viral fitness because of deficient

dNTP usage and that the E138K mutation can compensate for the deficit in dNTP usage associated with the M184I/V mutations [30]. The compensatory substitution at codon 230 in motif E contributes to reduced viral replication and has also been shown to confer resistance to all currently available NNRTIs in both phenotypic and biochemical assays [31].

Discussion

This study describes the MPS of proviral NFLGs and larger fragment from 26 well sampled groups of blood donors from PE who had previously been diagnosed as infected with subclade F1 (n = 25) and BF1 recombinant (n = 1) based on *pol* subgenomic fragment from cell free viruses using conventional bulk sequencing. The most remarkable observations in this study are that 23 of the 24 donors in whom genotyping was successful infected with HIV-1 BF1 recombinant variants. Of these, two novel BF1 CRFs with high genetic diversities that exceed >8% difference at both inter- and intra-host level were identified. These results suggest that both CRFs have been in circulation early in the epidemic and have been evolving independently ever since. Based on the similarity of their recombination profile, it is tempting to speculate

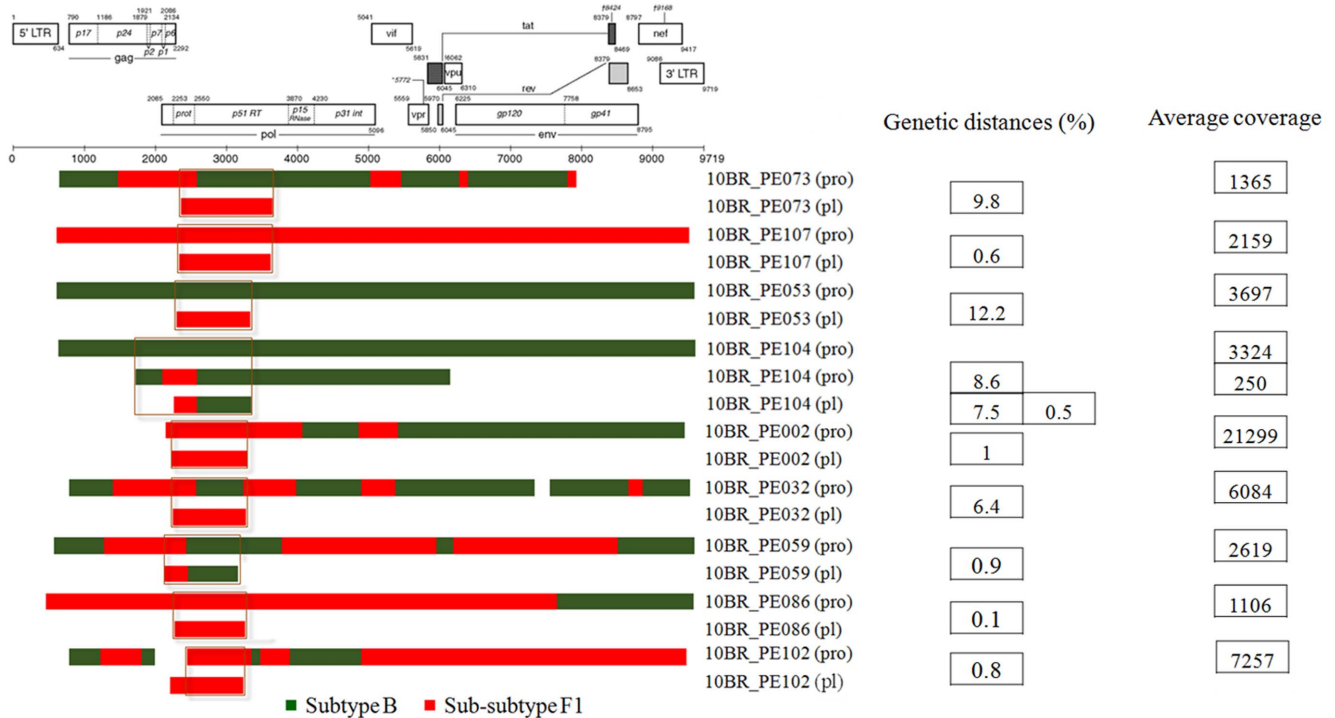


Figure 5. Schematic representation of the NFLGs, partial structure and breakpoint profiles of the BF1 sequences identified in this study from proviral DNA generated by deep sequencing approach and previously published cell free viruses generated by bulk sequencing approach. Sequences were mapped relative to the HXB2 numbering system. Genetic distances of overlapping regions (marked with orange boxes) between sequences from plasma and PBMCs together with the overall mean coverage depth are demonstrated. Distances were computed using the maximum composite likelihood method in MEGA version 6 [28]. As depicted in the figure, the intra-individual plasma and proviral sequence variation in four patients (10BR_PE073, 10BR_PE053, 10BR_PE104 and 10BR_PE032) in the partial *pol* regions (marked with orange boxes) were remarkably high. These results may indicate that the plasma viruses were derived from a population significantly distinct from those of the cellular sources; a result consistent with dual infection with different subtype. In sample 10BR_PE104, MPS data revealed the existence of subtype B NFLGs and a second BF1 recombinant strain (4450 bp) almost identical to the plasma virus in the same region. doi:10.1371/journal.pone.0112674.g005

that the CRF70_BF1 variants were old “second-generation” recombinants of CRF71_BF1 circulating in PE. Our estimates indicated that the CRF71_BF1 variants are responsible of 50% of infection caused by BF1 recombinants among blood donors in this region of northeast of Brazil. Moreover, if we assume no recombination in the remainder of the genome characterized as subtype B in the previous study, then the prevalence of BF1 recombinant variants is estimated at 20.4% (22/108) and CRF71_BF1 at 10.2% (11/108) of HIV-1 strains circulating among blood donors in PE. Given the high prevalence of CRF71_BF1 observed among the low risk blood donors, it may be suspected that a high prevalence for this variant could be found among other high-risk groups with tight transmission chain and that it has been able to break the transmission barrier from high-risk groups into the general population. Moreover, detection of both CRFs indicates that these variants are actively competing with other BF1 recombinants and other HIV-1 subtypes circulating in this region.

The proportion of BF1 variants described in this study is much greater than the rate reported in 169 individuals recently diagnosed as seropositive for HIV1 (2.8%) in the Metropolitan Region of Recife, Northeastern of Brazil [32] and in other study of 84 patients chronically infected with HIV (3.6%) naïve to antiretroviral treatment [33]. This difference is not surprising, because small fragments from different regions of HIV genomes were characterized in the previous studies while we used larger overlapped fragments to sequence the NFLGs, which is the

preferred method for accurate characterization of HIV-1 isolates. Thus, the previous studies are likely to have missed the recombinants samples. Despite the high rate of recombination as estimated using our fairly conservative assumptions, it is probable that our results have underestimated the true rate of infection with BF1 recombinant viruses; particularly our study was limited to donors infected with F1 and BF1 with partially sequenced viral fragments. Thus, it is possible that the BF1 infection in this group will be higher than what was observed if we had sequenced the virus NFLGs in all the 80 subtype B infected donors described in the previous study.

Additional observations of this study are the description of mixed infections with B, F1 and/or BF1 recombinants. The use of MPS technology enhanced our power to determine the dual infection of larger fragment of BF1 and genuine NFLGs of subtype B in the PBMCs of patient 10BR_PE104. The BF1 recombination profile was almost identical to the plasma virus in the same region demonstrating an intra-individual plasma and proviral sequence variation of only 0.5%. Thus, it is possible to assume that the primary infected PBMCs harboring the BF1 recombinants in this patient were likely the source of the plasma circulating viral sequences. The observation of an intact NFLGs of subtype B in the PBMCs of patient 10BR_PE104 may argue against the inactivity of this provirus but may agree with assumption that subtype B provirus may be integrated into the chromosome at a site at which its expression is prevented, or it may be transcriptionally inactive by virtue of being extrachromosomal [34]. It is also possible that

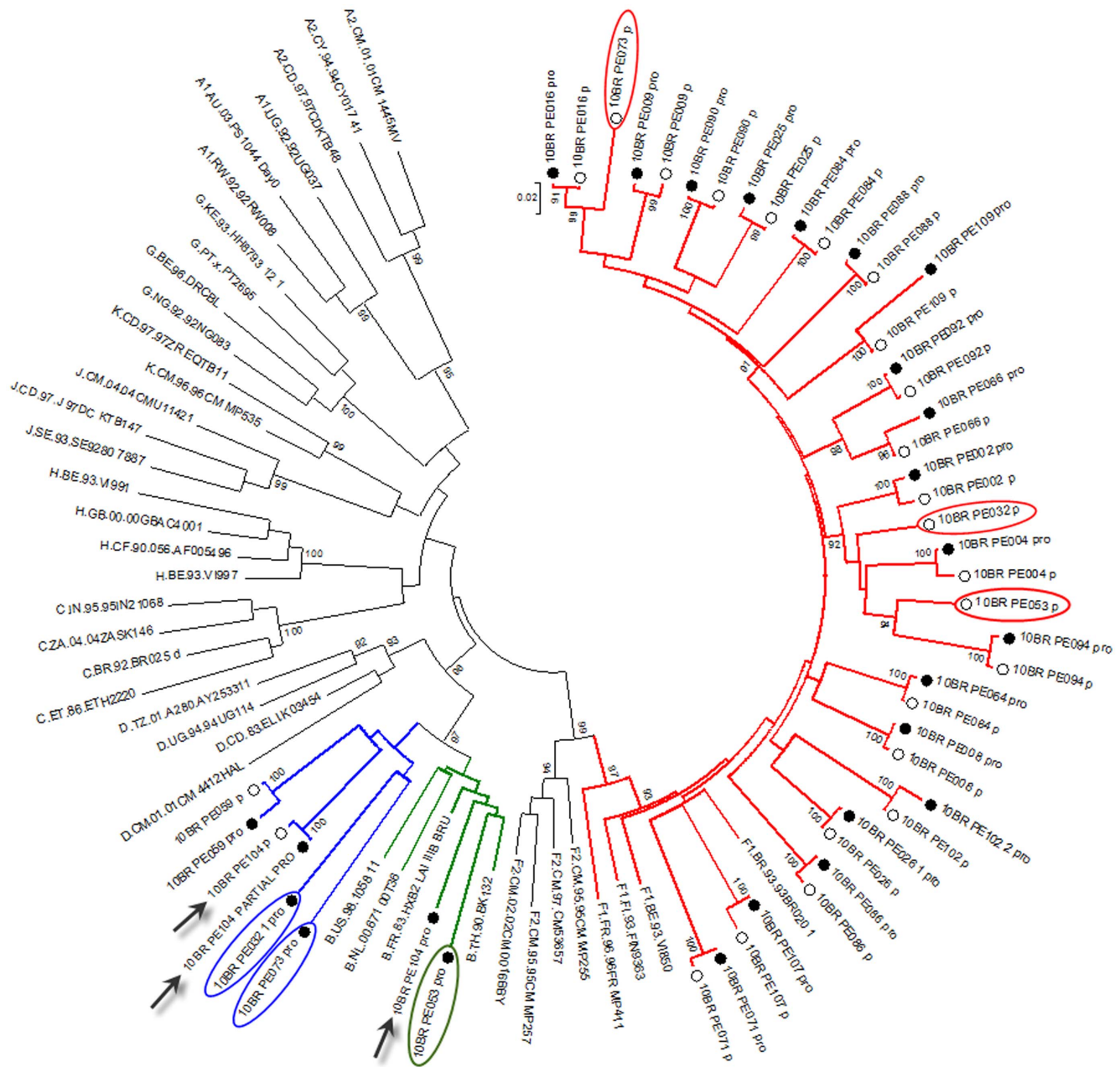


Figure 6. Comparison of phylogenetic clustering profile of sequences from both plasma (empty circles) and provirus (black circles) and other HIV-1 reference sequences from the Los Alamos HIV-1 database representing 11 genetic subtypes. Viral sequences from both compartments were aligned with the complete set of reference sequences obtained from the Los Alamos database (<http://hiv-web.lanl.gov>). Green, blue and red colored branches represent subtype B, BF1 recombinants, and subclade F1, respectively. Sequence with discordant results between PBMCs and plasma are marked with blue and red oval circles, respectively. For purposes of clarity, the tree was midpoint rooted. The approximate likelihood ratio test (aLRT) values of $\geq 90\%$ are indicated at nodes. The scale bar represents 0.02 nucleotide substitutions per site. Except subjects 10BR_PE073, 10BR_PE053, 10BR_PE104 and 10BR_PE032, each patient forms a tight cluster and is distinct from other subjects with aLRT SH-like supports $>95\%$ for all inter-subject clusters. The results from the ML analysis added further support to the results depicted in figure and were sufficiently robust to confirm the event of dual infection with different subtype. In case 10BR_PE104 (indicated by black arrow), MPS data revealed the existence of subtype B NFLGs and a second BF1 recombinant strain almost identical to the plasma virus in the same region. doi:10.1371/journal.pone.0112674.g006

the adaptation of the B strains to the PBMC in this patient allows them to avoid competitive exclusion by the dominant strain in plasma. In the other three samples with dual infections, subtype F1 seen in the plasma were completely absent in the PBMCs. Discordances in the HIV subtypes in both compartments may suggest differential sources of infecting viruses. It is also

conceivable that the discordances in the HIV subtypes in PBMCs and plasma are likely due to low-level minority strains present as B/BF1 variants that are not detected with bulk plasma sequencing or that the F1 viruses shed in the plasma were more fit. The evidence of dual infections in this study adds support to previous studies [35,36,37,38,39], as this event is far more common in

Table 2. Drug resistance mutations detected with bulk sequencing (plasma) and deep sequencing (PBMCs).

Sample ID	Protease gene coding region		Reverse transcriptase gene coding region (nRTIs)		Reverse transcriptase gene coding region (NNRTIs)	
	PLASMA (Bulk sequencing)	PBMC (Deep sequencing)	PLASMA (Bulk sequencing)	PBMC (Deep sequencing)	PLASMA (Bulk sequencing)	PBMC (Deep sequencing)
10BR_PE002*	M36I, V77I [†] , L89M	M36I, <u>V82I</u> , L89M		<i>M184I</i> *		<i>M230I</i> *
10BR_PE004*	M36I, V77I, L89M	M36I, <u>M46I</u> /V77I, L89M			V106I [†]	
10BR_PE008*	M36I, D60E, L89M	M36I, D60E, L89M				V106I
10BR_PE009*	L10V, M36I, L89M	L10V, M36I, L89M		<i>F77L</i> * [†] , <i>M184I</i> *		<i>M230I</i> *
10BR_PE016*	L10V, M36I, L89M	L10V, <i>G16E</i> , M36I, L89M				
10BR_PE025	L10V, K20R, M36I, L89M	L10V, K20R, M36I, L89M				
10BR_PE026	M36L, L89M	M36I, L89M				
10BR_PE032*	L10V, M36I, L89M	L10V, <u>L33F</u> *, M36I, L89M		<i>M184I</i> *		V108I, <i>M230I</i> *
10BR_PE053*	M36I [†] , L89M [†]	V77I				
10BR_PE059	L10I, M36I, L89M	L10I, M36I, L89M				
10BR_PE064	M36I, L89M	M36I, L89M				
10BR_PE066*	L10V, K20R [†] , M36I, D60E [†] , I62V, L63P, L89M, I93L	L10V, M36I, D60E, I62V, L89M, I93L				
10BR_PE071*	M36I	M36I		<i>F77L</i> * [†]		
10BR_PE073	M36I, I64M	M36I, I64M				
10BR_PE084	L10V, M36I, I64M, L89M	L10V, M36I, I64M, L89M				
10BR_PE086*	L10I [†] , K20R, M36I, I62V, L89M, I93L	K20R, M36I, I62V, L89M, I93L				
10BR_PE088	M36I, V82I, L89M	M36I, V82I, L89M			V106I	V106I
10BR_PE090*	L10I, M36I, L89M	L10I, M36I, <u>M46I</u> *, L89M		<i>M184I</i> *		
10BR_PE092	M36I, D60E, I62V, I64M, L89M	M36I, D60E, I62V, I64M, L89M				
10BR_PE094	K20R, M36I, L89M	K20R, M36I, L89M				
10BR_PE102*	L10I [†] , M36I [†] , I64L [†]					
10BR_PE104*	K20R [†] , M36I	M36I, <i>L63P</i> , <i>V77I</i> , <i>I93L</i>				
10BR_PE107*	M36I, I62V, L63P [†] , L89M	M36I, I62V, L89M			V108I [†]	
10BR_PE109*	K20M, M36I, D60E, L63P, L89M	K20M, M36I, <u>M46I</u> *, D60E, L63P, L89M		<i>F77L</i> * [†]		<i>E138K</i>

Samples displayed discordant genotypic data between the cell free viruses and the PBMC-viruses are marked by black dots, Major mutations are marked in boldface type, Major drug resistance mutations at prevalence >20% of the sequenced population are underlined, Major drug resistance mutations at prevalence <20% of the sequenced population are marked by star symbol, Mutations detect only in PBMCs are marked in italic face type.

[†]Transmitted drug resistance mutation, Mutations detect only in plasma are marked by pilcrow symbol.

doi:10.1371/journal.pone.0112674.t002

Brazil where these subtypes co-circulate. Furthermore, these data agree with the consensus that the presence of two or more HIV-1 subtypes within an infected individual is relatively frequent [40,41]. It is unclear from this study whether the occurrence of HIV multiple distinct strains was the result of superinfection with a second variant at a later time, or whether simultaneous infection with multiple viral strains occurred during a single transmission event. However, the circulation of multiple subtypes in Brazil fortifies the possibility of both scenarios. The overall results indicate that the rate of HIV-1 mixed infections within this Brazilian group is higher than 16%. From these results, we believed that the use of PBMC DNA in addition to plasma RNA is expected to provide highest sensitivity to detect mixed infection via population MPS. Whether dual infection and/or recombination had an impact on the clinical outcomes of the blood donors in this study was unknown, since the available clinical data resulted from one assessment that seek to understand risk exposures and motivations to donate blood.

Other important observation of this study is the underestimation of transmitted resistance obtained by routine plasma analysis that is revealed by the examination of the MPS populations of the archived proviruses in PBMCs. In this study, comparing both sources would have detected 11 DNA provirus disclosed DRMs by MPS not detected by routine plasma bulk analysis. In this study, eleven DNA provirus sequence had detectable DRMs previously missed by plasma bulk analysis. Six of these strains had 17 DRMs only in the PBMCs and four of which had detectable major DRMs at prevalence between 1–20% of the MPS population. These results support the previous observations that standard bulk sequencing cannot fully access the spectrum of viral variants archived in the proviral DNA [42]. The relatively higher proportion of recently infected donors carried low-abundance resistance mutations leads us to believe that the rate of the current transmitted DRMs is underestimated. It is well known that conventional Sanger sequencing of bulk PCR products are limited to the detection of high-frequency variants that present in greater

than 20% of the sequenced viral population [43,44]. However, we were able to detect additional DRMs at prevalence higher than 20% of the MPS provirus variants that had not been identified using plasma bulk sequencing. This result probably indicates that MPS approach permit characterization of considerable heterogeneity in the diversity and frequency variations in the proviral DNA [45]. Since the time of infection is unknown, it possible that some of the transmitted mutants in cell free viruses may disappear over time while persisting in cellular HIV-1 DNA [46,47,48]. All together, these results justify the inclusion of proviral DNA from PMBCs as a valuable source for resistance analysis, which is in agreement with previous reports [49,50]. On the other hand, we expected that our MPS approach would detect the entire mutations spectrum from the amplified and sequenced viral population that was displayed by conventional sequencing of the cell free viruses, but this does not hold true for some mutations. Seven patient samples with some detectable DRMs by bulk sequencing approach were not observed in PBMCs reads obtained through the MPS approach. One possible explanation beside the limiting factors intrinsic to PCR is that our MPS approach of the amplified fragments is not sufficiently sensitive to detect all DRMs. It is also possible that the discordant genotypes in these patients are due to dual infection. This was confirmed in patients 10BR_PE053 and 10BR_PE104 who displayed two phylogenetically distinct populations in the sequenced regions.

Apart from small sample size, the present study was limited by the investigation of only the HIV-1 proviruses. Although plasma HIV-1 RNA remains the material of choice for the determination of drug-resistant mutations and guiding therapeutic decisions [51,52] the proviral PBMC DNA sequence can contain a variety of multiple DRMs that are not present in plasma [47,53]. This, combined with the stability of DNA compared with RNA [54], and the fact that HIV DNA recovered from the proviral compartment can reliably be used for the determination of DRMs in treatment naïve patients [55] influenced our decision to use proviral DNA in this study.

Finally, needless to say, HIV phylogenetic analysis based on the complete genome sequences is more reliable than that of the *pol* or

other small partial fragment alone. The high percentage of the new CRF71_BF1 and other BF1 recombinants found among blood donors in Pernambuco, coupled with high rates of transmitted DRMs and dual infections confirm the need for effective surveillance to monitor the prevalence and distribution of HIV variants in a variety of settings in Brazil.

Acknowledgments

The Recipient Epidemiology and Donor Outcomes Study-III is the responsibility of the following persons:

Cesar de Almeida Neto¹, Alfredo Mendrone-Junior¹, Anna Bárbara Carneiro-Proietti², Divaldo de Almeida Sampaio³, Paula Loureiro³, Clarisse Lobo⁴ Maria Esther Lopes⁴, Ester Cerdeira Sabino⁵, Ligia Capuani⁵, João Eduardo Ferreira⁵, Pedro Losco Takecian⁵, Cláudia Di Lorenzo Oliveira⁶, Brian Custer⁷, Michael P. Busch^{7*}, Thelma T. Gonçalves⁷, Donald Brambilla⁸, Christopher McClure⁸, Simone A. Glynn⁹

¹Fundação Pró-Sangue Hemocentro de São Paulo, SP, Brazil.

²Fundação Hemominas, Belo Horizonte, MG, Brazil.

³Fundação Hemope, Recife, PE, Brazil.

⁴Hemorio, Rio de Janeiro, RJ, Brazil.

⁵University of São João Del-Rei, MG, Brazil

⁶University of São Paulo, SP, Brazil.

⁷Blood Systems Research Institute, San Francisco, CA, USA/University of California, San Francisco, San Francisco CA, USA

⁸RTI international

⁹National Heart, Lung, and Blood Institute, NIH

*Group leader:

Michael P. Busch, M.D., Ph.D.

Blood Systems Research Institute

270 Masonic Avenue

San Francisco, CA 94118-4417

mbusch@bloodsystems.org

Author Contributions

Conceived and designed the experiments: SSS. Performed the experiments: RP JTW PC SSS. Analyzed the data: RP JTW PC ACF SSS. Contributed reagents/materials/analysis tools: PL ECS MPB. Contributed to the writing of the manuscript: MPB SSS.

References

- Mansky LM, Temin HM (1995) Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 69: 5087–5094.
- Wei X, Ghosh SK, Taylor ME, Johnson VA, Emami EA, et al. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373: 117–122.
- Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387: 183–188.
- Sabino EC, Shpaer EG, Morgado MG, Korber BT, Diaz RS, et al. (1994) Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 68: 6340–6346.
- Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG (2003) Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrob Chemother* 51: 229–240.
- Rambaut A, Posada D, Crandall KA, Holmes EC (2004) The causes and consequences of HIV evolution. *Nat Rev Genet* 5: 52–61.
- Peeters M, Toure-Kane C, Nkengasong JN (2003) Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS* 17: 2547–2560.
- Thomson MM, Perez-Alvarez L, Najera R (2002) Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis* 2: 461–471.
- Konings FA, Haman GR, Xue Y, Urbanski MM, Hertzmark K, et al. (2006) Genetic analysis of HIV-1 strains in rural eastern Cameroon indicates the evolution of second-generation recombinants to circulating recombinant forms. *J Acquir Immune Defic Syndr* 42: 331–341.
- Arien KK, Vanham G, Arts EJ (2007) Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* 5: 141–151.
- Barreto CC, Nishyia A, Araujo LV, Ferreira JE, Busch MP, et al. (2006) Trends in antiretroviral drug resistance and clade distributions among HIV-1-infected blood donors in Sao Paulo, Brazil. *J Acquir Immune Defic Syndr* 41: 338–341.
- Sanabani S, Neto WK, de Sa Filho DJ, Diaz RS, Munerato P, et al. (2006) Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res Hum Retroviruses* 22: 171–176.
- Sanabani S, Kleine Neto W, Kalmar EM, Diaz RS, Janini LM, et al. (2006) Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. *Infect Genet Evol* 6: 368–377.
- Passaes CP, Guimaraes ML, Bello G, Morgado MG (2009) Near full-length genome characterization of HIV type 1 unique BC recombinant forms from Southern Brazil. *AIDS Res Hum Retroviruses* 25: 1339–1344.
- Guimaraes ML, Eyer-Silva WA, Couto-Fernandez JC, Morgado MG (2008) Identification of two new CRF_BF in Rio de Janeiro State, Brazil. *AIDS* 22: 433–435.
- Santos AF, Sousa TM, Soares EA, Sanabani S, Martinez AM, et al. (2006) Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* 20: 2011–2019.
- Alencar CS, Sabino EC, Carvalho SM, Leao SC, Carneiro-Proietti AB, et al. (2013) 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* 63: 387–392.
- Sanabani SS, Pastena ER, da Costa AC, Martinez VP, Kleine-Neto W, et al. (2011) Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in Sao Paulo, Brazil. *PLoS One* 6: e25869.
- Sanabani SS, Pastena ER, Neto WK, Martinez VP, Sabino EC (2010) Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in Sao Paulo, Brazil. *Virol J* 7: 74.

20. Sanabani SS, Pastena ER, Kleine Neto W, Barreto CC, Ferrari KT, et al. (2009) Near full-length genome analysis of low prevalent human immunodeficiency virus type 1 subclade F1 in Sao Paulo, Brazil. *Virol J* 6: 78.
21. Thomson MM, Sierra M, Tanuri A, May S, Casado G, et al. (2004) 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* 20: 1126–1133.
22. Pessoa R, Watanabe JT, Nukui Y, Pereira J, Kasseb J, et al. (2014) Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by illumina massively parallel sequencing technology. *PLoS One* 9: e93374.
23. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066.
24. Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 11: 1423–1425.
25. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.
26. Schultz AK, Zhang M, Leitner T, Kuiken C, Korber B, et al. (2006) A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes. *BMC Bioinformatics* 7: 265.
27. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267.
28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729.
29. Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, et al. (1999) Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* 13: 2349–2359.
30. Xu HT, Oliveira M, Quashie PK, McCallum M, Han Y, et al. (2012) Subunit-selective mutational analysis and tissue culture evaluations of the interactions of the E138K and M184I mutations in HIV-1 reverse transcriptase. *J Virol* 86: 8422–8431.
31. Xu HT, Quan Y, Schader SM, Oliveira M, Bar-Magen T, et al. (2010) The M230L nonnucleoside reverse transcriptase inhibitor resistance mutation in HIV-1 reverse transcriptase impairs enzymatic function and viral replicative capacity. *Antimicrob Agents Chemother* 54: 2401–2408.
32. Cavalcanti AM, Brito AM, Salustiano DM, Lima KO, Silva SP, et al. (2012) Recent HIV infection rates among HIV positive patients seeking voluntary counseling and testing centers in the metropolitan region of Recife - PE, Brazil. *Braz J Infect Dis* 16: 157–163.
33. de Medeiros LB, Lacerda HR, Cavalcanti AM, de Albuquerque Mde F (2006) Primary resistance of human immunodeficiency virus type 1 in a reference center in Recife, Pernambuco, Brazil. *Mem Inst Oswaldo Cruz* 101: 845–849.
34. Simmonds P, Balfe P, Peutherer JF, Ludlam CA, Bishop JO, et al. (1990) Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J Virol* 64: 864–872.
35. Soares de Oliveira AC, Pessoa de Farias R, da Costa AC, Sauer MM, Bassichetto KC, et al. (2012) Frequency of subtype B and F1 dual infection in HIV-1 positive, Brazilian men who have sex with men. *Virol J* 9: 223.
36. Sanabani SS, Pessoa R, Soares de Oliveira AC, Martinez VP, Giret MT, et al. (2013) Variability of HIV-1 genomes among children and adolescents from Sao Paulo, Brazil. *PLoS One* 8: e62552.
37. Ramos A, Tanuri A, Schechter M, Rayfield MA, Hu DJ, et al. (1999) Dual and recombinant infections: an integral part of the HIV-1 epidemic in Brazil. *Emerg Infect Dis* 5: 65–74.
38. Diaz RS, Sabino EC, Mayer A, Mosley JW, Busch MP (1995) Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group. *J Virol* 69: 3273–3281.
39. Blackard JT, Mayer KH (2004) HIV superinfection in the era of increased sexual risk-taking. *Sex Transm Dis* 31: 201–204.
40. Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, et al. (2012) The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J Infect Dis* 206: 267–274.
41. Cornelissen M, Pasternak AO, Grijzen ML, Zorgdrager F, Bakker M, et al. (2012) HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. *Clin Infect Dis* 54: 539–547.
42. Wirlden M, Soulie C, Valantin MA, Fourati S, Simon A, et al. (2011) Historical HIV-RNA resistance test results are more informative than proviral DNA genotyping in cases of suppressed or residual viraemia. *J Antimicrob Chemother* 66: 709–712.
43. Van Laethem K, Van Vaerenbergh K, Schmit JC, Sprecher S, Hermans P, et al. (1999) Phenotypic assays and sequencing are less sensitive than point mutation assays for detection of resistance in mixed HIV-1 genotypic populations. *J Acquir Immune Defic Syndr* 22: 107–118.
44. Gunthard HF, Wong JK, Ignacio CC, Havlir DV, Richman DD (1998) Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. *AIDS Res Hum Retroviruses* 14: 869–876.
45. Bansode V, McCormack GP, Crampin AC, Ngwira B, Shrestha RK, et al. (2013) Characterizing the emergence and persistence of drug resistant mutations in HIV-1 subtype C infections using 454 ultra deep pyrosequencing. *BMC Infect Dis* 13: 52.
46. Geretti AM, Fox Z, Johnson JA, Booth C, Lipscomb J, et al. (2013) Sensitive assessment of the virologic outcomes of stopping and restarting non-nucleoside reverse transcriptase inhibitor-based antiretroviral therapy. *PLoS One* 8: e69266.
47. Bon I, Gibellini D, Borderi M, Alessandrini F, Vitone F, et al. (2007) Genotypic resistance in plasma and peripheral blood lymphocytes in a group of naive HIV-1 patients. *J Clin Virol* 38: 313–320.
48. Parisi SG, Boldrin C, Cruciani M, Nicolini G, Cerbaro I, et al. (2007) Both human immunodeficiency virus cellular DNA sequencing and plasma RNA sequencing are useful for detection of drug resistance mutations in blood samples from antiretroviral-drug-naive patients. *J Clin Microbiol* 45: 1783–1788.
49. Jakobsen MR, Tolstrup M, Sogaard OS, Jorgensen LB, Gorry PR, et al. (2010) Transmission of HIV-1 drug-resistant variants: prevalence and effect on treatment outcome. *Clin Infect Dis* 50: 566–573.
50. Bon I, Alessandrini F, Borderi M, Gorini R, Re MC (2007) Analysis of HIV-1 drug-resistant variants in plasma and peripheral blood mononuclear cells from untreated individuals: implications for clinical management. *New Microbiol* 30: 313–317.
51. Bi X, Gatanaga H, Ida S, Tsuchiya K, Matsuoka-Aizawa S, et al. (2003) Emergence of protease inhibitor resistance-associated mutations in plasma HIV-1 precedes that in proviruses of peripheral blood mononuclear cells by more than a year. *J Acquir Immune Defic Syndr* 34: 1–6.
52. Kaye S, Comber E, Tenant-Flowers M, Loveday C (1995) The appearance of drug resistance-associated point mutations in HIV type 1 plasma RNA precedes their appearance in proviral DNA. *AIDS Res Hum Retroviruses* 11: 1221–1225.
53. Riva E, Pistello M, Narciso P, D'Offizi G, Isola P, et al. (2001) Decay of HIV type 1 DNA and development of drug-resistant mutants in patients with primary HIV type 1 infection receiving highly active antiretroviral therapy. *AIDS Res Hum Retroviruses* 17: 1599–1604.
54. Banks L, Gholamin S, White E, Zijenah L, Katzenstein DA (2012) Comparing Peripheral Blood Mononuclear Cell DNA and Circulating Plasma viral RNA pol Genotypes of Subtype C HIV-1. *J AIDS Clin Res* 3: 141–147.
55. Chew CB, Potter SJ, Wang B, Wang YM, Shaw CO, et al. (2005) Assessment of drug resistance mutations in plasma and peripheral blood mononuclear cells at different plasma viral loads in patients receiving HAART. *J Clin Virol* 33: 206–216.