Title: Molecular and geographic characterization of hiv-1 bf recombinant viruses

Authors: Juliana Sacramento M Souza, Jailton J Silva Júnior, Carlos Brites, Joana P Monteiro-Cunha

PII: S0168-1702(19)30268-0
DOI: https://doi.org/10.1016/j.virusres.2019.197650
Article Number: 197650

Reference: VIRUS 197650

To appear in: *Virus Research*

Received date: 17 April 2019
Revised date: 1 July 2019
Accepted date: 3 July 2019

Please cite this article as: { https://doi.org/

MOLECULAR AND GEOGRAPHIC CHARACTERIZATION OF HIV-1 BF RECOMBINANT VIRUSES

Authors: Juliana Sacramento M Souza[a], Jailton J Silva Júnior[a], Carlos Brites[b], Joana P Monteiro-Cunha[a]

Author's affiliation:

[a]Universidade Federal da Bahia, Instituto de Ciências da Saúde, Departamento de Bioquímica e Biofísica, Salvador, Bahia, Brazil

[b]Universidade Federal da Bahia, Complexo Hospitalar Prof. Edgard Santos, Laboratório de Pesquisa em Infectologia, Salvador, BA, Brazil.

Corresponding author:

Joana P. Monteiro-Cunha

Núcleo de Bioinformática, Departamento de Bioquímica e Biofísica, Universidade Federal da Bahia. Av. Reitor Miguel Calmon, s/n, Salvador, Brasil # 55 71 3283-8914 - Fax. # 55 71 3283-8894. E-mail: joana.monteiro@ufba.br

Highlights

- Several BF recombination patterns found in 16 countries; most from Brazil
- BF recombinant viruses circulate in Brazil since at least 1985
- Highest recombination frequencies in *pol*, *gag* and *env* genes; lowest in *tat* and *rev*
- Predominance of subtype B in most accessory and regulatory genes, except in *vif*
- There are preferential recombination points within the HIV genome
- Most of BF recombinant viruses use CCR5 as a coreceptor for viral entry
- X4/dual tropism is less frequent among F subtype sequences, compared with B subtype

Abstract

The Human Immunodeficiency Virus Type 1 (HIV-1) presents a wide genetic variability, which is represented by four groups, nine subtypes of group M and several

recombinant forms. Among these, the BF recombinants have been distinguished by a high global dispersion and an increase in number and diversity. To date, 15 BF Circulating Recombinant Forms (CRFs) and diverse BF Unique Recombinant Forms (URFs) have been described. In Brazil, nine CRF_BF have been identified. The aim of this work was to perform molecular and geographic characterization of HIV-1 BF recombinant strains. Near full-length genomes of 265 BF recombinant viruses were collected from public databases and molecular analyses were performed. These sequences were originally retrieved between 1993-2006 and isolated from 16 countries (51.3% from Brazil). Diagnostic's year analysis showed that BF recombinants circulate in Brazil since at least 1985. Most sequences displayed recombination in the *pol* (84.9%), *gag* (69.3%) and *env* (51.4%) regions. The subtype B predominated in all accessory and regulatory genes, except in *vif*, in which the F subtype was predominant (40.4%). Twelve regions with a recombination rate higher than 10% were identified, especially one region inside p24 gene (1359-1397) whose recombination was present in more than 30% of the sequences. Coreceptor usage prediction during viral entry showed that BF recombinants preferentially use CCR5 (67.2%) and the most frequent tetrapeptides found in the V3 loop were GPGR (47.9%) and GPGQ (21.1%). The frequency of X4/dual viruses was lower amongst F subtype (25.8%) V3 sequences, compared with B subtype (43%). In addition, mutations associated with intermediate or high resistance levels to PI (10.6%), NRTI (15.0%), NNRTI (14.0%) and INSTI (2.6%) were identified. The great diversity of the recombination patterns evidences that the recombination between the subtypes B and F is frequent, reflecting a probable high rate of dual infection and the acquisition of advantageous characteristics for viral fitness.

Keywords: HIV; BF; Recombination; CRF; genome.

1. Introduction

The HIV-1 is characterized by a great genetic diversity which is attributed to the high error and recombination rates of the reverse transcription process, as well as to the fast virions turnover in infected individuals (Santoro and Perno, 2013). This high HIV-1 strain variability is represented by four groups (M, N, O, and P), which are associated with independent transmission events. The M group is further classified into nine subtypes (A, B, C, D, F, G, H, J, and K) and many recombinants forms. Some of these

recombinant viruses become fixed and undergo rapid expansion in the population, and they are classified as Circulating Recombinant Forms (CRFs). To date, 98 HIV-1 CRFs have been described in many countries (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html) and are estimated to be responsible for 18-20% of infections worldwide (Hemelaar et al., 2006). Among those, 15 CRFs (15.3%) have been originated from recombination between subtypes B and F, which is the most frequent dual recombination among HIV-1 subtypes (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html).

The first CRF_BF was identified in Argentina in 2001 (Carr et al., 2001). Since then, these BF recombinants have been described in several countries, mainly in Brazil, where nine CRFs_BF and several BF URFs have been identified. The importance of BF recombinants in the HIV-1 epidemic has been demonstrated by several studies and reinforced by the increasing number of these variants among the infected population (De Oliveira et al., 2008; Villanova, 2010; Reis et al., 2017). Previous studies have reported 10% to 30% prevalence of BF recombinants in Brazil (Sanabani et al., 2013; Pessôa et al., 2016; da Costa et al., 2016; Reis et al., 2017) . In these studies, the F1 subtype frequencies were always lower than those for BF recombinants. The high diversity of circulating BF strains may be related to the acquisition of adaptive advantages considering that recombinant viruses must overcome the parental strains in the coinfected individual in order to be efficiently transmitted (Melo et al., 2012). In fact, different lines of evidence have demonstrated that BF recombinants may exhibit distinct biological characteristics. For instance, a rapid growth rate of CRF12_BF and CRF38_BF recombinants in Argentina and Uruguay has been observed (Aulicino et al., 2007; Bello et al., 2010). In another study, the *vpu* genomic region of BF recombinants was associated with an increased production of viral particles in cell assays when compared with subtype B (De Candia et al., 2010).

The aim of this work was to broadly and comprehensively characterize the genetic diversity of BF recombinant genomes circulating worldwide. We mapped the origin of these isolates and investigated molecular characteristics, such as recombination patterns, the occurrence of preferential recombination points within the HIV genome, the frequency of parental subtypes (B and F) fragments within the recombinant full sequences, the drug resistance profile and the prediction of membrane coreceptors usage during the virus entry into the host cell. This information may contribute to a better understanding of HIV epidemiology and the viral characteristics that could be

associated with genetic evolution. Furthermore, the knowledge of HIV genetic variation represents a critical point for the development of drugs and effective vaccines.

2. Materials and methods

2.1. Data collection

Two hundred and sixty-five Near Full-Length Genome (NFLG) sequences (> 7000 bp) of HIV-1 BF recombinants viruses were collected from the HIV sequencing database maintained by Los Alamos National Laboratory (LANL) and NCBI Nucleotide database until September 2018. The information from these two databases were cross-referenced along with data available in the original studies that have originated those sequences. When available, the data were retrieved, as following: GenBank accession number; patient identification in LANL; sequence size; article in which the sequence has been described; subtype reported by NCBI, LANL and by the original article; country and city of sample collection; gender and age of the patient; TCD4, TCD8 lymphocytes and viral load counts; transmission mode; risk group; year of infection, diagnosis and sample collection; clinical condition of the patient; coreceptor usage and viral phenotype; and HIV treatment. The sequences were identified with the GenBank accession number, country abbreviation, and sample collection year (i.e., KT427758.BR.2010).

2.2. Alignment

The 265 recombinant sequences were aligned with 39 reference sequences of all HIV-1 M subtypes, collected in the LANL database (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) and group O sequence (GenBank: L20587.1). The sequence alignment was performed using MAFFT version 7 default parameters (mafft --thread 10 --threadtb 5 --threadit 0 --reorder --auto input > output) (Katoh et al., 2017).

2.3. Recombination and phylogenetics analysis

Recombination patterns were evaluated using jpHMM online tool (http://jphmm.gobics.de/submission_hiv.html). The intervals for recombination breakpoints based on HXB2 numbering obtained for each sequence were used to identify the regions where the recombination between B and F subtypes were more frequent.

The best-fit model for phylogenetic analysis was chosen using the IQ-TREE model selection tool (http://iqtree.cibiv.univie.ac.at/). According to corrected Akaike Information Criteria (AICc) the best-fit model was GTR+F+I+G4. The reconstruction of the evolutionary history of these sequences was estimated using a Maximum likelihood (ML) analysis using the software IQ-TREE (Command: iqtree -s aln.fasta -m GTR+F+I+G4 -bb 10000 -alrt 10000 -nt AUTO) (Trifinopoulos et al., 2016). The tree visualization and editing were performed using FigTree v 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). The genetic similarity between the recombinant sequences, as well as, the parental subtypes were inferred from the mean pairwise distances in non-overlapping windows of 100 nucleotides across the alignment. Distances were measured with the Kimura two-parameter model implemented in MEGA X (Kumar et al., 2018) using 500 bootstrap replicates to calculate the standard deviations.

2.4. Prediction of coreceptor usage

The viral tropism prediction was determined according to the amino acid composition of the V3 region of the HIV-1 envelope gp120 protein, using the Geno2pheno tool (http://coreceptor.geno2pheno.org/index.php) with a false-positivity rate (FPR) cutoff of 20% (Lengauer et al., 2007). The tetrapeptide motifs in the V3 loop were identified through manual search in the alignment.

2.5. Drug resistance analysis

HIV-1 drug resistance was determined throughout analyses of the reverse transcriptase, protease, and integrase regions in the *pol* gene. The identification of drug resistance-associated mutations was performed using the Stanford University Genotypic Resistance Interpretation Algorithm program (HIVdb v.8.7) (https://hivdb.stanford.edu/hivdb/by-sequences/).

2.6. Statistical analysis

The GraphPad Prism Version 7 (GraphPad Software Inc., CA, USA) was used to perform the statistical analyses, based on Fisher's exact (two-tailed) and Kruskal-Wallis tests. P-values ≤ 0.05 were considered significant.

3. Results

3.1. Sample characterization

The 265 NFLG of HIV BF recombinant viruses were collected from NCBI and LANL databases (**S1**). These sequences were originally obtained between 1993 and 2016 (median: 2005). The mean age of the infected individuals (available for n=103) was 33 years at the time of collection. The majority (66%) were in the age group of 20 to 39 years, which coincides with the age group with the highest prevalence of HIV infections worldwide (www.who.int/hiv/en/). The sequences (median size 8923 bp, ranging from 7293-9770 bp) were obtained from 16 different countries and most of them (51.3%) were from Brazil (**Fig. 1A**). Sixty-one per cent of the sequences were classified as URFs_BF (161/265and the remaining as CRFs_BF (104/265, 39.2%) (**Fig. 1B**).

3.2. Recombination and phylogenetics analysis

Fig.2 shows the recombination patterns of the 15 previously identified CRF_BF **(Fig.2)** and all URF_BF (https://tinyurl.com/BFjpHMM) generated using jpHMM. The structural, accessory and regulatory genomic regions were individually analyzed for the occurrence of intragenic recombination. The *pol* gene (84.9%) was the genomic region in which recombination occurred more frequently, followed by *gag* (69.3%) and *env* (51.4%) regions. Although 40.4% of the genomes have maintained sequences of the F subtype in the *vif* gene, the B subtype sequences were predominant in all other regulatory and accessory genes (**Fig. 3A**).

The frequency at which each nucleotide position (based on HXB2 numbering) appears within a recombination region was calculated based on the recombination breakpoint intervals provided by the jpHMM analysis. For this purpose, all recombination intervals

of each distinct recombinant pattern were cataloged in a spreadsheet. Then, the frequency of each individual nucleotide position within these recombinant regions was calculated. We identified 12 genomic regions where recombination occurred in at least 10% of the analyzed sequences (**Fig. 3B).** It was also possible to identify regions in which recombination is more frequently found. A recombination frequency higher than 30% (1359-1397 bp) was found in p24 gene. Besides that, recombination is present in 20-25% of the sequences in three regions (p24: 1400-1418; protease: 2463-2476; reverse transcriptase: 3691-3728). In order to evaluate the possible association between breakpoints locations and intersubtype genetic conservation, the genetic distances (0.04 – 0.22) within nucleotide windows containing the regions in which recombination occurred most frequently were compared with those (0.05 – 0.11) within windows covering regions with low recombination frequencies. No statistical differences were observed between the mean distances in these regions (Kruskal-Wallis test, $p>0.05$). Reconstruction of the evolutionary history of the sequences is represented in the ML tree (**S2**). Collapsed groups of all CRF_BF and correlated sequences (**Fig. 4)** showed statistical support for all except the CRF12 cluster. The reference groups that are not the parental subtypes also showed statistical support.

3.3. GP120 analysis

The prediction of the coreceptor usage indicated that 67.2% (178/265) of the sequences were from viruses that use CCR5 as coreceptor to enter the cell while 32.8% (87/265) were classified as X4 or dual tropic (R5/X4) viruses. Only four sequences showed recombination inside the GP120 V3 loop region while in six V3 sequences the subtype could not be assigned. Among the V3 sequences classified as F, 74.2% (115/155) were predicted as R5 while among the V3 sequences of subtype B 57% (57/100) were R5 viruses (**Fig. 5A)**. Twenty-six different tetrapeptide motifs in the V3 loop were found: GPGR (47.92%), GPGQ (21.13%), GWGR (6.03%), RPRR (0.37%), RPGR (0.75%), GFGR (1.88%), GWGK (0.75%), GLGR (0.75%), APGR (6.79%), GRGR (1.13%), GPGG (1.50%), GPGS (0.75%), GPGK (4.90%), [AG]PG[RS] (0.37%), [AG]PGR (0.75%), GTGQ (0.37%), GPGH (0.37%), GPER (0.37%), G[PA]G[RS] (0.37%), GSGR (0.37%), GGVG (0.37%), RPGQ (0.37%), AWGR (0.37%), KPGR (0.37%), TPGR (0.37%), GWGQ (0.37%). The subtype (**Fig 5B)** and tropism (**Fig 5C)** associated with the five most frequent motifs were also investigated.

3.4. Antiretroviral (ARV) resistance

Despite the absence of information regarding the use of ARV by the patients from whom the sequences were collected, the presence of mutations associated with therapeutic resistance was investigated among the BF genomes. Mutations conferring resistance to at least one antiretroviral class were found in 114 sequences (43.0%), regardless of the associated level of resistance (**Table 1**). Out of these, 67 (25.3%) displayed mutations associated with resistance to one ARV class, 30 (11.3%) to two classes, 14 (5.3%) to three classes and 3 (1.1%) to all four ARV classes. Thirty-five (13.2%) sequences presented mutations associated with some level of resistance to protease inhibitors (PIs) (**Fig. 6A**). Seventy-four (29%) sequences showed at least one resistance mutation within the reverse transcriptase region. Out of these, 46 (17.4%) had mutations that were associated with some level of resistance to nucleoside reverse transcriptase inhibitors (NRTIs) (**Fig. 6B**) and 59 (22.3%) to the non-nucleoside reverse transcriptase inhibitors (NNRTIs) (**Fig. 6C**). In the integrase region, 21 different drug resistance-associated mutations were identified among the sequences. Among these mutations, 10 were classified as major and only seven (2.6%) sequences showed mutations associated with intermediate or high resistance level to INSTIs (**Fig. 6D)**.

4. Discussion

In the present study, 265 NFLG from BF recombinant viruses originated in 16 different countries were collected. This shows the global dispersion of these recombinants that were initially described in South America (Aulicino et al., 2007) (**Fig 1**). Among the individuals, the earliest diagnosis occurred in 1985. The viral sequence of this Brazilian patient was obtained in 2005 and classified as CRF29. Therefore, BF recombinant strains were most likely circulating even before the description of the first recombinant in Brazil in 1992 (Melo et al., 2012). Nearly 51% of the recombinant samples were isolated in Brazil and BF recombinants with Brazilian parental viruses were previously identified in Italy and Japan. These observations point to the important role of BF recombinants originated in Brazil in the global HIV/AIDS pandemic (Bruselles et al., 2009).

The variety of recombination patterns between subtypes B and F (**Fig.2**) indicates that these recombination events are frequent and lead to the emergence of viable and transmissible viral strains, further enhancing the HIV-1 genetic diversity. In some cases, strains with different recombinant profiles have exhibited similar mosaic structures, but with unique patterns in part of their genomes, suggesting their generation by successive recombination processes from a common recombinant ancestor. Several factors may contribute to the high emergence frequency and diversity of recombinant strains, such as co-circulation of multiple HIV-1 subtypes in the same region, and co- or super-infection of individuals with multiple subtypes. (Thomson et al., 2002; Leal et al., 2008). Notably, HIV-1 recombinant viruses are most prevalent in areas where multiple subtypes co-circulate. In Brazil, both B (since around mid to late 1960s) and F (since around late 1970s) subtypes have co-circulated for a long period of time (Guimara et al., 2007). Moreover, BF recombinants have grown both in number of CRFs and URFs diversity (Monteiro et al., 2009; Pessoa et al., 2014). Therefore, the wide variety of recombination patterns found in our study reflects a probable high rate of dual infection (coinfection and superinfection) between subtypes B and F. These findings led us to believe in an urgent need for public health measures to prevent these episodes, especially because the generation of viral diversity associated with recombination is a challenge for epidemic control. Beyond that, the detection of BF recombinant viruses in the population may be currently underestimated, since most subtyping studies are commonly based only on one genomic fragment.

The likelihood of new biological properties in recombinant progenies may occur due multiple recombination profiles, conferring them adaptive advantages over the parental subtypes. A previous study on the *in vitro* HIV recombination consequences have observed that an increase in the proportion of recombinant genomes over time coincides with the progressive loss of one of the parental strains (Iglesias-Sanchez and Lopez-Galindez, 2002). In fact, it has been suggested that the increasing prevalence of BF viruses in Brazil, reaching 35% in some regions (Guimarães et al., 2015), may be related to the disappearance of pure subtype F, which is usually found at low frequency (Sanabani et al., 2006). Likewise, the HIV epidemic in some other countries or regions, such as Malaysia and West Africa, is undergoing a gradual replacement of the initial predominant HIV-1 pure subtype strains by CRFs (Lau and Wong, 2013).

Due to recombination in the viral population, genetic variability may enhance the chance of adaptation of the recombinant progeny upon changes in selection pressures in

the host. As a result of HIV-1 recombination, the emergence of viral strains that escape the host immune response (Streeck et al., 2008) and/or develop resistance to antiretrovirals has been demonstrated (Kellam and Larder, 1995; Moutouh et al., 1996). Therefore, it is possible that the crossover points do not occur randomly and that there are genomic regions favoring recombination (Minin et al., 2007). The highest recombination rates were observed inside *pol, gag,* and *env* regions **(Fig. 3A)**. The high rate of recombination within these structural genes have been widely reported (Dykes et al., 2004; Galetto et al., 2004; Galetto et al., 2006; Monteiro-Cunha et al., 2011). In particular, the high recombination rate in *pol* has been associated with its composition by homopolymer bases, sequence similarity and the presence of secondary structures, favoring the template switching during reverse transcription (Jetzt et al., 2000; Onafuwa-Nuga and Telesnitsky, 2009; Sanjuán and Domingo-Calap, 2016). On the other hand, the regions comprising the first exon of *rev* (*rev1)* and the second exon of *tat* (*tat2*, the major form) showed the lowest frequencies of recombination (0% and 1.2% respectively). Similarly, the *vpu* region preferentially maintained the pure B subtype (60.8%) in the BF recombinant genomes (**Fig. 3A**), although it has been previously shown that a structurally reorganized BF recombinant *vpu* has an increased ability to enhance viral replication in vitro when compared to its parental B (De Candia et al., 2010).

vif is an important viral infectivity factor that prevents the action of cellular proteins of the APOBEC-3G family and was the only region in which the F subtype was preferentially maintained (40.4%), In all other accessory/regulatory regions, the subtype B has predominated. It has been shown that most of the *vif* alleles classified as subtype F were highly effective in neutralizing the A3H hapII protein, whereas the alleles of other non-B subtypes had low activity for the same protein (Binka et al., 2012). Similarly, a recent study showed that natural *vif* proteins of F1 subtypes efficiently attack hA3G and hA3H hapII proteins, even though they were unable to degrade hA3C and hA3F (Zhang et al., 2016). Therefore, the maintenance of F subtype in the *vif* region may confer some advantage to these recombinants and should be better investigated by functional assays. In agreement with our findings, a recent study showed evidence of non-random inheritance of genomic fragments from different HIV-1 M subtypes among 283 distinct recombination profiles (Tongo et al., 2018). Taken together, these results show differences in the frequency with which HIV-1 subtypes are maintained in the

different genomic regions of the recombinant sequences, which may be related to the maintenance of advantageous characteristics for viral fitness.

Moreover, in the present study, 12 distinct recombination regions were identified in more than 10% of the sequences and were found within genes like *p17*, *p24, protease, p51* (RT), *p15, p31, vif, vpr, vpu/gp120, gp41/rev2* and *nef/3'LTR* **(Fig. 3B)**. In particular, the region 1359-1397 (relative to HXB2 reference) found in p24 capsid protein gene (aa 59-71) showed the highest frequency of recombination (>30%). Although it has been previously suggested that genetic similarity is a factor favoring recombination (An and Telesnitsky, 2002), our results do not point to a direct relationship between genetic conservation and distribution of recombination hotspots, indicating that factors other than genetic similarity should influence the occurrence of preferential regions for recombination in the HIV-1 genome. In fact, the study by Jia and colleagues has suggested that the probability of base pairing along the viral RNA secondary structure may be a more important factor influencing recombination than sequence conservation (Jia et al., 2016).

Genetic variations within the V3 domain of the gp120 envelope glycoprotein are associated with coreceptor and cell line tropism (Jiang et al., 2010). Viruses using the CCR5 coreceptor are more common in early infections and have higher transmission rates than X4 or R5/X4 viruses which are associated with more rapid disease progression (Cuevas et al., 2010). Thus, the dispersion of BF recombinants may be related to the maintenance of the R5 phenotype in the majority (67.2%) of these viruses. From this result, it can also be estimated that most patients infected with BF recombinants would respond to Maraviroc, the CCR5 coreceptor inhibitory drug. HIV-1 viruses of other subtypes can use both coreceptors. However, it has been shown that subtype D viruses are mostly dual-tropic and that subtypes C and G have lower X4 virus rates than subtype B (Huang et al., 2007). In the present analysis, the frequency of X4/dual tropic viruses was lower amongst F subtype (25.8%) V3 sequences, compared with B subtype (43%) (Fisher's test, p=0.006) (**Fig. 5A**). All 26 structural motifs present among the BF sequences have been previously described. Sequences with the GPGR, GPGQ and APGR motifs are mostly classified as subtype F at the V3 loop region and those with the GWGR and GPGK as B subtype **(Fig. 5B)**. Indeed, the GWGR motif is known as a characteristic domain of Brazilian subtype B and has been previously associated with the exclusive use of the CCR5 receptor (Leal et al., 2008). However, in the present study, five (31.3%) out of 16 sequences with the GWGR motif were

predicted as R5/X4. Similarly, other studies have reported that approximately 25% of viruses with this structural motif can use the CXCR4 receptor (X4 or R5X4) (Ferraro et al., 2001; Villanova, 2010; Sanabani et al., 2011). All sequences with the APGR motif were predicted as CCR5 virus **(Fig. 5C)**.

The drug resistance analysis showed that among the PI drugs, DRV/r (97.7%) and NFV (86.8%) showed the highest and lowest susceptibility levels respectively (**Fig. 6A**). In Brazil, ATV/r or DRV/r are the recommended PI after the first therapeutic failure. On the other side, LPV/r is being excluded from the main international protocols because of the high toxicity and side effects. Among the NRTIs, TDF (89.8%) presented the highest susceptibility level while DDI (17%), ABC (15%), AZT (14%) and D4T (14%) showed the highest resistance levels **(Fig. 6B)**. Within the NNRTIs class, 13.2% of the sequences presented high or intermediate resistance level to NVP and 11.7% to EFV **(Fig. 6C)**. Until 2017, EFV combined with the TDF and 3TC constituted the first-line antiretroviral therapy in Brazil, but since 2018, EFV was replaced by integrase inhibitor DTG in the Brazilian guidelines. Of note, only 2.6% (7/265) of BF sequences were associated with high/intermediate level of resistance to INSTIs **(Fig. 6D)**. Most of mutations in this region are accessories and among them, G163R (8.3%) and G163K (4.2%) were the most frequent (**Table 1**). These mutations, have been reported at high frequency among BF recombinants (Delicado et al., 2016). Noteworthy, only 0.4% of the sequences was associated with intermediate-level resistance to DTG. This drug is now recommended as a first- line antiretroviral drug in the United States, Brazil and Europe (http://aidsinfo.nih.gov). Therefore, this results corroborate with previous studies that reported the high genetic barrier to resistance as one of the advantages of DTG (Walmsley et al., 2013; Raffi et al., 2015).

Nonetheless, the real impact of the selective pressure exerted by each individual drug on the emergence of drug resistance mutations could not be better evaluated in the present study, due to the absence of treatment information of the BF carriers. Despite this, all sequences that showed mutations associated with resistance to PIs, NRTIs and NNRTIs were retrieved after at least one of the drugs associated with a given mutation had already been FDA approved. In contrast, the sequences with mutations associated with high/intermediate level INSTI resistance have been collected before (between 1999 and 2006) the FDA approval of these drugs. Although these mutations have been previously associated with the use of INSTIs (Wensing et., 2017; https://hivdb.stanford.edu/dr-summary/resistance-notes/INSTI/), they probably represent natural variation among the

sequences analyzed in this study. In order to evaluate the natural occurrence of INSTIs associated mutations, we searched all available sequences in the Los Alamos database collected before 2007, when the first Integrase inhibitor (Raltegravir) was started. Among the 2996 sequences collected until 2006, 47 (1.6%) presented mutations associated with high or intermediate resistance to INSTIs. There is no statistical difference between this result and the one found in our study (Fisher's test, p=0.2), indicating indeed that these mutations occur naturally in the viral population. In addition, these data are in line with the results of a study by Delicado et al, 2016, showing high prevalence of INSTI-associated mutations among F subtype and BF recombinant sequences isolated from drug naive subjects. Taken together, these findings suggest that some of the INSTI-associated mutations could be subtype-associated polymorphisms and that their role in the development of antiretroviral resistance should be individually analyzed for each genetic variant of HIV-1.

In conclusion, this report showed the circulation of HIV-1 BF recombinant strains worldwide. Notably, NFLG of these variants were not detected in Africa. Most of these viruses circulate in Brazil and although nine CRFs_BF have been characterized in this country, these CRFs did not represent the majority of the sequences found. On the contrary, a great diversity of URFs was identified. There are genomic regions in which recombination occurs more frequently and regions that preferentially maintain sequences of pure subtype B. *vif* was the only genomic region that presented higher frequency of pure F subtype, which could be associated with possible adaptive advantages. Considering that HIV genotypic variations reflect in several aspects such as efficacy of diagnostic tests, disease monitoring, immune response, treatment outcome and vaccine design, the high genetic variability observed among the genomic structures strongly highlights the importance of monitoring the HIV molecular epidemiology for better control of the epidemic.

Conflict of interest

There is no conflict of interest among coauthors.

Funding

Fundação de Amparo à Pesquisa do Estado da Bahia.

References

An, W., Telesnitsky, A., 2002. Effects of Varying Sequence Similarity on the Frequency of Repeat Deletion during Reverse Transcription of a Human Immunodeficiency Virus Type 1 Vector 76, 7897–7902. https://doi.org/10.1128/JVI.76.15.7897

Aulicino, P.C., Holmes, E.C., Rocco, C., Mangano, A., Sen, L., 2007. Extremely Rapid Spread of Human Immunodeficiency Virus Type 1 BF Recombinants in Argentina. J. Virol. 81, 427–429. https://doi.org/10.1128/JVI.01403-06

Bello, G., Aulicino, P.C., Ruchansky, D., Guimarães, M.L., Lopez-Galindez, C., Casado, C., Chiparelli, H., Rocco, C., Mangano, A., Sen, L., Morgado, M.G., 2010. Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. Retrovirology 7, 22. https://doi.org/10.1186/1742-4690-7-22

Binka, M., Ooms, M., Steward, M., Simon, V., 2012. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. J. Virol. 86, 49–59. https://doi.org/10.1128/JVI.06082-11

Bruselles, A., Rozera, G., Bartolini, B., Prosperi, M., Del Nonno, F., Narciso, P., Capobianchi, M.R., Abbate, I., 2009. Use of Massive Parallel Pyrosequencing for Near Full-Length Characterization of a Unique HIV Type 1 BF Recombinant Associated with a Fatal Primary Infection. AIDS Res. Hum. Retroviruses 25, 937–942. https://doi.org/10.1089/aid.2009.0083

Carr, J.K., Avila, M., Gomez Carrillo, M., Salomon, H., Hierholzer, J., Watanaveeradej, V., Pando, M. a, Negrete, M., Russell, K.L., Sanchez, J., Birx, D.L., Andrade, R., Vinoles, J., McCutchan, F.E., 2001. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. AIDS 15, F41-7. https://doi.org/10.1097/00002030-200110190-00002

Cuevas, M.T., Fernández-García, A., Pinilla, M., García-Álvarez, V., Thomson, M., Delgado, E., González-Galeano, M., Miralles, C., Serrano-Bengoechea, E., Ojea de Castro, R., López-Álvarez, M.J., Lezáun, M.J., Sánchez-García, A.M., Sánchez-Martínez, M., Muñoz-Nieto, M., Pérez-Álvarez, L., 2010. Short Communication: Biological and Genetic Characterization of HIV Type 1 Subtype B and Nonsubtype B Transmitted Viruses: Usefulness for Vaccine Candidate Assessment. AIDS Res. Hum. Retroviruses 26, 1019–1025. https://doi.org/10.1089/aid.2010.0018

da Costa, C.M., Costa de Oliveira, C.M., Chehuan de Melo, Y.F., Delatorre, E., Bello, G., Couto-Fernandez, J.C., 2016. High HIV-1 Genetic Diversity in Patients from Northern Brazil. AIDS Res. Hum. Retroviruses 00, aid.2016.0044. https://doi.org/10.1089/aid.2016.0044

De Candia, C., Espada, C., Duette, G., Ghiglione, Y., Turk, G., Salomón, H., Carobene, M., 2010. Viral replication is enhanced by an HIV-1 intersubtype recombination-derived Vpu protein. Virol. J. 7, 259. https://doi.org/10.1186/1743-422X-7-259

De Oliveira, C.M., Almeida, F.J., Rodrigues, R., Crozatti, M., Vazquez, C.M.P., Do Socorro Carneiro Ferrão, M., Campeas, A.E., Marques, S.R., Berezin, E.N., De Macedo Brígido, L.F., 2008. High frequency of BF mosaic genomes among HIV-1-infected children from Sao Paulo, Brazil. Arch. Virol. 153, 1799–1806. https://doi.org/10.1007/s00705-008-0178-7

Delicado, T.L., Torrecilla, E., Holguín, África, 2016. Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naive individuals. J. Antimicrob. Chemother. 71, 362–366. https://doi.org/10.1093/jac/dkv333

Dykes, C., Balakrishnan, M., Planelles, V., Zhu, Y., Bambara, R.A., Demeter, L.M., 2004. Identification of a preferred region for recombination and mutation in HIV-1 gag. Virology 326, 262–279. https://doi.org/10.1016/j.virol.2004.02.033

Ferraro, G.A., Antônio, M., Mello, G., Sutmöller, F., 2001. Biological Characterization and Chemokine Receptor Usage of HIV Type 1 Isolates Prevalent in Brazil 17, 1241–1247.

Galetto, R., Giacomoni, V., Véron, M., Negroni, M., 2006. Dissection of a circumscribed recombination hot spot in HIV-1 after a single infectious cycle. J. Biol. Chem. 281, 2711–

2720. https://doi.org/10.1074/jbc.M505457200
Galetto, R., Moumen, A., Giacomoni, V., Véron, M., Charneau, P., Negroni, M., 2004. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. J. Biol. Chem. 279, 36625–36632. https://doi.org/10.1074/jbc.M405476200
Guimara, M.L., Bello, G., Eyer-silva, W.A., Chequer-fernandez, S.L., Teixeira, S.L.M., Morgado, M.G., 2007. Demographicl history of HIV-1 subtypes B and F in Brazil 7, 263–270. https://doi.org/10.1016/j.meegid.2006.11.002
Guimarães, M.L., Marques, B.C.L., Bertoni, N., Teixeira, S.L.M., Morgado, M.G., Bastos, F.I., De Brito, A.M., Koller, E.M.P., Dourado, M.I.C., Vettorazzi, M.L.T., Ralil, M., Santos, N.T.V., Paes, P.C.D., Leone, S., Andrade, T., Moyses, S.T., Querrer, V.P.S., Caiaffa, W.T., 2015. Assessing the HIV-1 epidemic in Brazilian drug users: A molecular epidemiology approach. PLoS One 10, 1–17. https://doi.org/10.1371/journal.pone.0141372
Iglesias-Sánchez, M.J., López-Galíndez, C., 2002. Analysis, Quantification and Evolutionary Consequences of HIV-1 in Vitro Recombination. Virology 304, 392–402. https://doi.org/10.1006/viro.2002.1657
Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20, W13–W23. https://doi.org/10.1097/01.aids.0000247564.73009.bc
Huang, W., Eshleman, S.H., Toma, J., Fransen, S., Stawiski, E., Paxinos, E.E., Whitcomb, J.M., Young, A.M., Donnell, D., Mmiro, F., Musoke, P., Guay, L.A., Jackson, J.B., Parkin, N.T., Petropoulos, C.J., 2007. Coreceptor Tropism in Human Immunodeficiency Virus Type 1 Subtype D: High Prevalence of CXCR4 Tropism and Heterogeneous Composition of Viral Populations. J. Virol. 81, 7885–7893. https://doi.org/10.1128/JVI.00218-07
Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., Dougherty, J.P., 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol. 74, 1234–40. https://doi.org/10.1128/JVI.74.3.1234-1240.2000.Updated
Jia, L., Li, L., Gui, T., Liu, S., Li, H., Han, J., Guo, W., Liu, Y., Li, J., 2016. Analysis of HIV-1 intersubtype recombination breakpoints suggests region with high pairing probability may be a more fundamental factor than sequence similarity affecting HIV-1 recombination. Virol. J. 13, 1–12. https://doi.org/10.1186/s12985-016-0616-1
Jiang, X., Burke, V., Totrov, M., Williams, C., Cardozo, T., Gorny, M.K., Zolla-Pazner, S., Kong, X.-P.P., 2010. Conserved structural elements in the V3 crown of HIV-1 gp120. Nat. Struct. Mol. Biol. 17, 955–961. https://doi.org/10.1038/nsmb.1861
Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 1–7. https://doi.org/10.1093/bib/bbx108
Kellam, P., Larder, B. a, 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J. Virol. 69, 669–674.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K.,2018. MEGA X: Molecular Evolutionary. Genetics Analysis across Computing Platforms. Mol Biol Evol. 35,1547-1549. https://doi.org/10.1093/molbev/msy096
Lau, K.A., Wong, J.J.L., 2013. Current trends of HIV recombination worldwide. Infect. Dis. Rep. 5, 15–20. https://doi.org/10.4081/idr.2013.s1.e4
Leal, É., Silva, W.P., Sucupira, M.C., Janini, L.M., Diaz, R.S., 2008. Molecular and structural characterization of HIV-1 subtype B Brazilian isolates with GWGR tetramer at the tip of the V3-loop. Virology 381, 222–229. https://doi.org/10.1016/j.virol.2008.08.029
Lengauer, T., Sander, O., Sierra, S., Thielen, A., Kaiser, R., 2007. Bioinformatics prediction of HIV coreceptor usage. Nat. Biotechnol. 25, 1407–1410. https://doi.org/10.1038/nbt1371
Melo, F.L., Jamal, L.F., Zanotto, P.M.D.A., 2012. Characterization of primary isolates of HIV type 1 CRF28_BF, CRF29_BF, and unique BF recombinants circulating in São Paulo, Brazil. AIDS Res. Hum. Retroviruses 28, 1082–8. https://doi.org/10.1089/AID.2011.0123
Minin, V.N., Dorman, K.S., Fang, F., Suchard, M.A., 2007. Phylogenetic mapping of recombination hotspots in human immunodeficiency virus via spatially smoothed change-point processes. Genetics 175, 1773–1785. https://doi.org/10.1534/genetics.106.066258

Ministério da Saúde do Brasil, 2017. Hiv, aids. HIV Aids Bol. Epidemiológico 64.

Monteiro-Cunha, J.P., Araujo, A.F., Santos, E., Galvao-Castro, B., Alcantara, L.C.J., 2011. Lack of High-Level Resistance Mutations in HIV Type 1 BF Recombinant Strains Circulating in Northeast Brazil. AIDS Res. Hum. Retroviruses 27, 623–631. https://doi.org/10.1089/aid.2010.0126

Monteiro, J.P., Alcantara, L.C.J., de Oliveira, T., Oliveira, A.M., Melo, M.A.G., Brites, C., Galvão-Castro, B., 2009. Genetic variability of human immunodeficiency virus-1 in Bahia state, Northeast, Brazil: High diversity of HIV genotypes. J. Med. Virol. 81, 391–399. https://doi.org/10.1002/jmv.21414

Moutouh, L., Corbeil, J., Richman, D.D., 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. Proc. Natl. Acad. Sci. 93, 6106–6111. https://doi.org/10.1073/pnas.93.12.6106

Onafuwa-Nuga, A., Telesnitsky, A., 2009. The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination. Microbiol. Mol. Biol. Rev. 73, 451–480. https://doi.org/10.1128/MMBR.00012-09

Pessoa, R., Carneiro Proietti, A.B.D.F., Busch, M.P., Sanabani, S.S., 2014. Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil. Genome Announc. 2, e00386-14-e00386-14. https://doi.org/10.1128/genomeA.00386-14

Pessôa, R., Loureiro, P., Esther Lopes, M., Carneiro-Proietti, A.B.F., Sabino, E.C., Busch, M.P., Sanabani, S.S., 2016. Ultra-Deep Sequencing of HIV-1 near Full-Length and Partial Proviral Genomes Reveals High Genetic Diversity among Brazilian Blood Donors. PLoS One 11, e0152499. https://doi.org/10.1371/journal.pone.0152499

Raffi, F., Rachlis, A., Brinson, C., Arasteh, K., Górgolas, M., Brennan, C., Pappa, K., Almond, S., Granier, C., Nichols, W.G., Cuffe, R.L., Jr, J.E., Walmsley, S., 2015. Dolutegravir efficacy at 48 weeks in key subgroups of treatment-naive HIV-infected individuals in three randomized trials. AIDS 29, 167–174. https://doi.org/10.1097/QAD.0000000000000519

Reis, M.N.D.G., Bello, G., Guimarães, M.L., Stefani, M.M.A., 2017. Characterization of HIV-1 CRF90_BF1 and putative novel CRFs_BF1 in Central West, North and Northeast Brazilian regions. PLoS One 12, 1–16. https://doi.org/10.1371/journal.pone.0178578

Sanabani, S., Neto, W.K., Kalmar, E.M.N.N., Diaz, R.S., Janini, L.M., Sabino, E.C., 2006. Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in São Paulo, Brazil. Infect. Genet. Evol. 6, 368–377. https://doi.org/10.1016/j.meegid.2006.01.003

Sanabani, S.S., Pastena, É.R. de S., da Costa, A.C., Martinez, V.P., Kleine-Neto, W., de Oliveira, A.C.S., Sauer, M.M., Bassichetto, K.C., Oliveira, S.M.S., Tomiyama, H.T.I., Sabino, E.C., Kallas, E.G., 2011. Characterization of Partial and Near Full-Length Genomes of HIV-1 Strains Sampled from Recently Infected Individuals in São Paulo, Brazil. PLoS One 6, e25869. https://doi.org/10.1371/journal.pone.0025869

Sanabani, S.S., Pessôa, R., Soares de Oliveira, A.C., Martinez, V.P., Giret, M.T.M., de Menezes Succi, R.C., Carvalho, K., Tomiyama, C.S., Nixon, D.F., Sabino, E.C., Kallas, E.G., 2013. Variability of HIV-1 Genomes among Children and Adolescents from São Paulo, Brazil. PLoS One 8, e62552. https://doi.org/10.1371/journal.pone.0062552

Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. Cell. Mol. Life Sci. 73, 4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Santoro, M.M., Perno, C.F., 2013. HIV-1 Genetic Variability and Clinical Implications. ISRN Microbiol. 2013, 1–20. https://doi.org/10.1155/2013/481314

Streeck, H., Li, B., Poon, A.F.Y., Schneidewind, A., Gladden, A.D., Power, K.A., Daskalakis, D., Bazner, S., Zuniga, R., Brander, C., Rosenberg, E.S., Frost, S.D.W., Altfeld, M., Allen, T.M., 2008. Immune-driven recombination and loss of control after HIV superinfection. J. Exp. Med. 205, 1789–1796. https://doi.org/10.1084/jem.20080281

Thomson., M.M., Delgado, E., Herrero, I., Villahermosa, M.L., Vázquez-de Parga, E., Cuevas, M.T., Carmona, R., Medrano, L., Pérez-Alvarez, L., Cuevas, L., Nájera, R., 2002. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near

full-length genome sequences. J Gen Virol 83, 107-19. https://doi.org/doi:10.1099/0022-1317-83-1-107

Tongo, M., de Oliveira, T., Martin, D.P., 2018. Patterns of genomic site inheritance in HIV-1M inter-subtype recombinants delineate the most likely genomic sites of subtype-specific adaptation. Virus Evol. 4, 1–9. https://doi.org/10.1093/ve/vey015

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44, W232–W235. https://doi.org/10.1093/nar/gkw256

Villanova, F.E., 2010. Diversity of HIV-1 Subtype B : Implications to the Origin of BF Recombinants 5, 1–9. https://doi.org/10.1371/Citation

Walmsley, S.L., Antela, A., Clumeck, N., Duiculescu, D., Eberhard, A., Gutiérrez, F., Hocqueloux, L., Maggiolo, F., Sandkovsky, U., Granier, C., Pappa, K., Wynne, B., Min, S., Nichols, G., 2013. Dolutegravir plus Abacavir–Lamivudine for the Treatment of HIV-1 Infection. N. Engl. J. Med. 369, 1807–1818. https://doi.org/10.1056/NEJMoa1215541

Wensing, A.M., Calvez, V., Günthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., Shafer, R.W., Richman, D.D., 2017. Update of the Drug Resistance Mutations in HIV-1. Top Antivir Med. 24, 132-133.

Zhang, Z., Gu, Q., Jaguva, A., Jeyaraj, M., Schmidt, S., Zielonka, J., Heckel, J., Cichutek, K., Häussinger, D., Smits, S.H.J., 2016. Vif Proteins from Diverse Human Immunodeficiency Virus / Simian Immunodeficiency Virus Lineages Have Distinct Binding Sites in A3C. J. Virol. 90, 10193–10208. https://doi.org/10.1128/JVI.01497-16.Editor

FIGURE CAPTIONS

Fig. 1. Distribution of 265 HIV BF recombinant viruses (**A)** Geographic origin and worldwide distribution of the strains. (N) number of collected sequences in each country. The map shows the estimated number of people living with HIV in 2016 by WHO region. Figure adaptation from http://gamapserver.who.int/mapLibrary/Files/Maps/HIV_all_2016.png. **(B)** Genetic profile distribution among BF recombinant viruses circulating in the world. URF_BF: BF Unique Recombinant Forms. CRF: Circulating Recombinant Form.

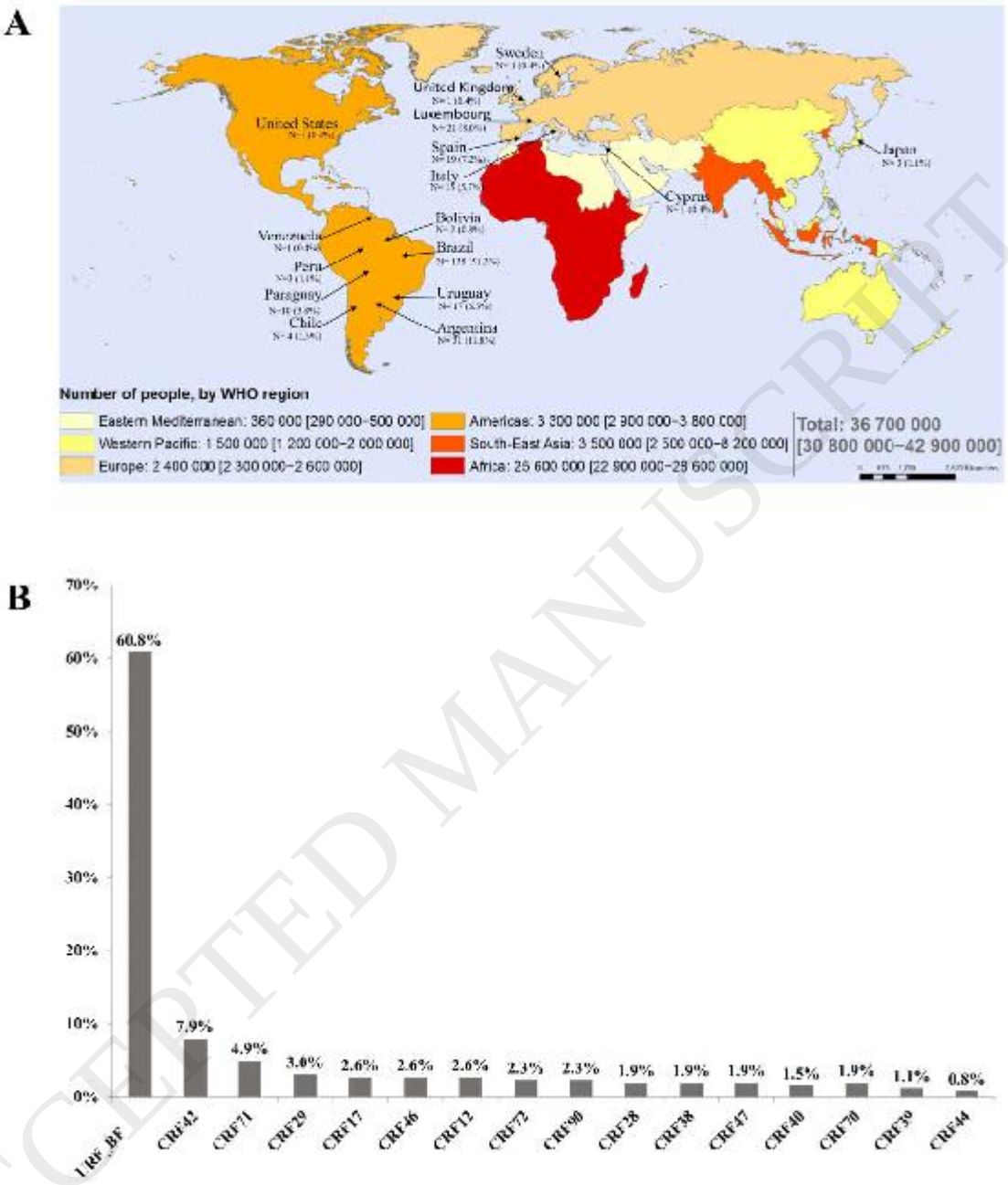

Fig. 2. Recombination patterns of the 15 CRF_BF circulating worldwide. The CRFs described in Brazil are identified in bold. Adapted from Los Alamos HIV sequence database (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html).

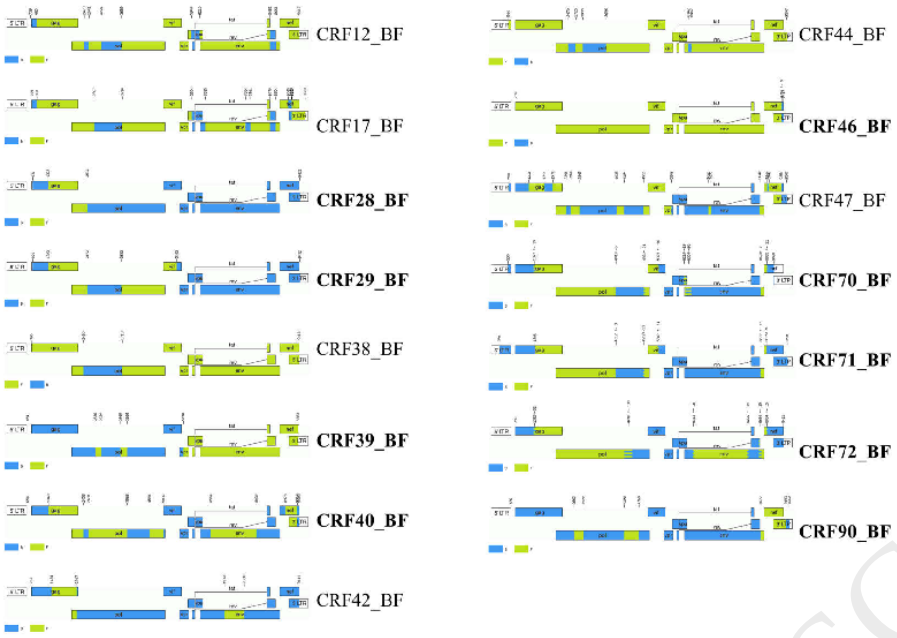

Fig. 3. Preferential recombination points within the HIV genome **(A)** Parental subtype and recombination frequencies within each HIV genomic region among BF recombinant viruses. (**B)** Nucleotide positions involved in recombination events present in more than 10% of the analyzed sequences. Twelve regions are identified with their respective nucleotide positions in the graph and in the genomic map.

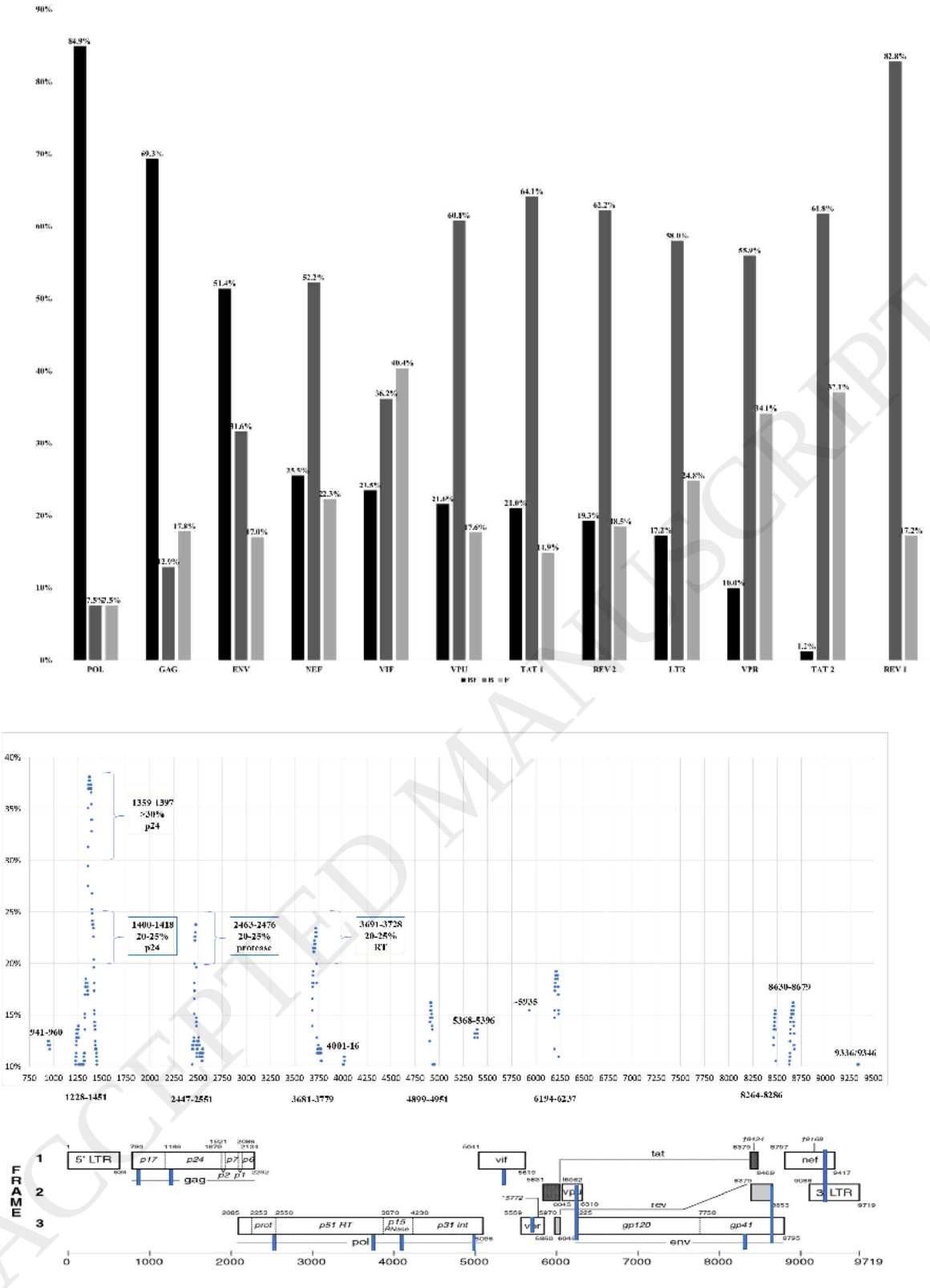

Fig. 4. ML tree midpoint rooted showing the phylogenetic relationship among the HIV-1 BF recombinant viruses. The scale represents the substitution of 0.2 nucleotides per site. Each color represents one CRF_BF and related sequences that clustered in the same clade. Sequences in black represent URF_BF that did not clustered with any CRF_BF.

Statistically supported clades are identified by ***** if SH-aLRT>=80% and Ultrafast bootstrap (UF-Boot) >=95%.

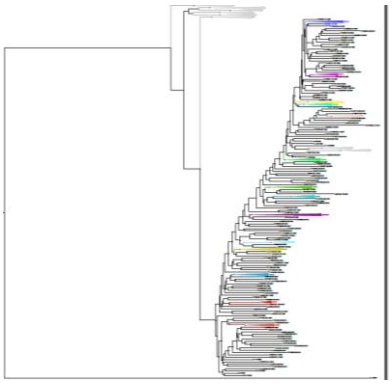

Fig. 5 Association between coreceptor usage, HIV-1 subtype and amino acid composition of the gp120 V3 region. (A) Frequency of R5 and X4/Dual tropic strains among subtypes B and F V3 sequences. (B) Classification of sequences with the most frequent motifs in the GP120 V3 loop according to the sequence subtype and (C) viral tropism.

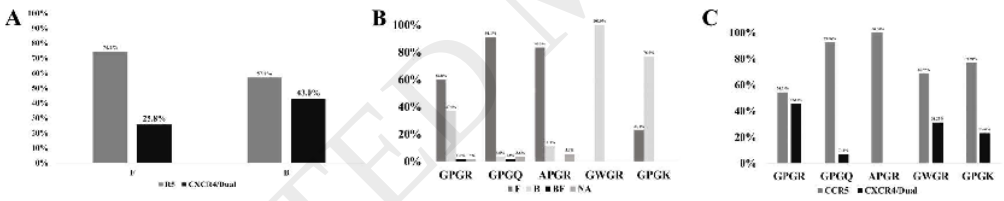

Fig. 6. Prevalence of antiretroviral resistance among HIV-1 BF recombinant sequences. Resistance to **(A)** Protease Inhibitor (PIs). **(B)** Nucleoside Reverse Transcriptase Inhibitor (NRTIs). **(C)** Non-Nucleoside Reverse Transcriptase Inhibitor (NNRTIs). **(D)** Integrase Strand Transfer Inhibitors (INSTIs). Protease Inhibitors: ATV (Atazanavir/ritonavir), DRV/r (Darunavir/ritonavir), FPV (Fosamprenavir/ritonavir), IDV/r (Indinavir/ritonavir), LPV/r (Lopinavir/ritonavir), NFV (Nelfinavir), SQV (Saquinavir/ritonavir) and TPV/r (Tipranavir/ritonavir). Nucleoside Reverse Transcriptase Inhibitors: ABC (Abavavir), AZT (Zidovudine), D4T (Stavudine), DDI (Didanosine), FTC (Emtricitabin), 3TC (Lamivudine) and TDF (Tenofovir). Non-Nucleoside Reverse Transcriptase Inhibitors: DOR (Doravirine), EFV (Efavirenz), ETR

(Etravirine), NVP (Nevirapine) and RPV (Rilpivirin). Integrase Inhibitors: BIC (Bictegravir), DTG (Dolutegravir), EVG (Elvitegravir) and RAL (Raltegravir).

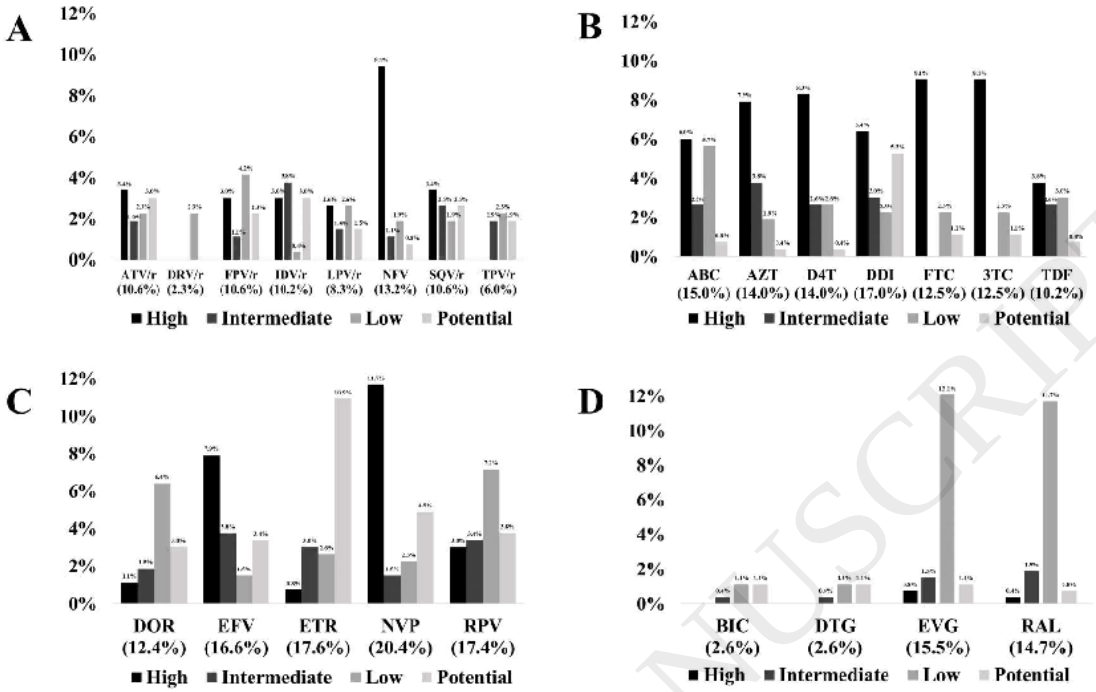

Table 1 – Drug resistance mutations among 265 BF recombinant sequences.

NRTI			NNRTI			PI						INSTI					
						Major			Accessory			Major			Accessory		
Mutation	**n**	**%**	**Mutation**	**n**	**%**	**Mutation**	**n**	**%**	**Mutation**	**n**	**%**	**Mutation**	**n**	**%**	**Mutation**	**n**	**%**
M41L	21	7.92	**K103N**	16	6.04	**L90M**	11	.15	**N88D**	8	3.02	**E138EK**	2	0.75	**G163R**	22	8.30
M184V	19	7.17	**E138A**	9	3.40	**D30N**	9	.40	**K20T**	7	2.64	**R263RK**	2	0.75	**G163K**	11	4.15
T215Y	19	7.17	**G190A**	7	2.64	**V82A**	8	.02	**L33F**	5	1.89	**G140S**	1	0.38	**S153Y**	1	0.38
D67N	16	6.04	**Y181C**	6	2.26	**I54V**	7	.64	**L10F**	3	1.13	**T66TP**	1	0.38	**S153SF**	1	0.38
K70R	14	5.28	**V108I**	6	2.26	**M46I**	5	.89	**K43T**	2	0.75	**P145PS**	1	0.38	**G163KR**	1	0.38
L210W	10	3.77	**N348I**	4	1.51	**M46MI**	3	.13	**G48R**	2	0.75	**N155D**	1	0.38	**G140R**	1	0.38
K219Q	8	3.02	**V179D**	4	1.51	**L76V**	3	.13	**F53L**	2	0.75	**Y143YH**	1	0.38	**T97TA**	1	0.38
T215F	8	3.02	**A98G**	3	1.13	**M46L**	2	.75	**Q58E**	1	0.38	**E138K**	1	0.38	**G163EK**	1	0.38
L74V	5	1.89	**H221Y**	3	1.13	**I84V**	2	.75	**L24I**	1	0.38	**E92K**	1	0.38	**G140Deletion**	1	0.38
E44D	4	1.51	**K238T**	3	1.13	**V82F**	1	.38	**G73S**	1	0.38	**N155Deletion**	1	0.38	**A128T**	1	0.38
M184I	3	1.13	**M230I**	3	1.13	**I50L**	1	.38	**G48GR**	1	0.38				**T97A**	1	0.38
T69D	3	1.13	**E138K**	2	0.75	**D30DN**	1	.38	**G48GE**	1	0.38						
D67G	2	0.75	**K103S**	2	0.75	**V82IT**	1	.38	**G48E**	1	0.38						
E40F	2	0.75	**L100I**	2	0.75												
E44A	2	0.75	**V106A**	2	0.75												
L74I	2	0.75	**E138EK**	1	0.38												
T215S	2	0.75	**F227L**	1	0.38												
V75A	2	0.75	**G190GA**	1	0.38												
V75M	2	0.75	**G190GEKR**	1	0.38												
F77L	1	0.38	**K101E**	1	0.38												
K219E	1	0.38	**K101PQ**	1	0.38												
K219R	1	0.38	**K103E**	1	0.38												
M184MI	1	0.38	**K103KE**	1	0.38												
M184MV	1	0.38	**K103KI**	1	0.38												
T215D	1	0.38	**K103KN**	1	0.38												
T215E	1	0.38	**P236L**	1	0.38												
T215TNSY	1	0.38	**V179E**	1	0.38												
V75T	1	0.38															

n= number of sequences with mutation.