



# Estimating HIV-1 Genetic Diversity in Brazil Through Next-Generation Sequencing

Brunna M. Alves<sup>1</sup>, Juliana D. Siqueira<sup>1</sup>, Isabel M. Prellwitz<sup>1</sup>, Ornella M. Botelho<sup>1</sup>, Vanusa P. Da Hora<sup>2</sup>, Sabri Sanabani<sup>3</sup>, Patrícia Recordon-Pinson<sup>4</sup>, Hervé Fleury<sup>4</sup>, Esmeralda A. Soares<sup>1</sup> and Marcelo A. Soares<sup>1,5\*</sup>

<sup>1</sup> Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, Brazil, <sup>2</sup> Laboratório de Biologia Molecular, Escola de Medicina, Universidade Federal do Rio Grande, Rio Grande do Sul, Brazil, <sup>3</sup> LIM-3, Hospital das Clínicas FMUSP, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, <sup>4</sup> CNRS MFP-UMR 5234, University Hospital of Bordeaux, University of Bordeaux, Bordeaux, France, <sup>5</sup> Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

## OPEN ACCESS

### Edited by:

Joris Hemelaar,  
University of Oxford, United Kingdom

### Reviewed by:

Frank van der Meer,  
University of Calgary, Canada  
Guido van Marle,  
University of Calgary, Canada

### \*Correspondence:

Marcelo A. Soares  
masoares@inca.gov.br

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 15 October 2018

**Accepted:** 25 March 2019

**Published:** 09 April 2019

### Citation:

Alves BM, Siqueira JD, Prellwitz IM, Botelho OM, Da Hora VP, Sanabani S, Recordon-Pinson P, Fleury H, Soares EA and Soares MA (2019) Estimating HIV-1 Genetic Diversity in Brazil Through Next-Generation Sequencing. *Front. Microbiol.* 10:749. doi: 10.3389/fmicb.2019.00749

Approximately 36.7 million people were living with the human immunodeficiency virus (HIV) at the end of 2016 according to UNAIDS, representing a global prevalence rate of 0.8%. In Brazil, an HIV prevalence of 0.24% has been estimated, which represents approximately 830,000 individuals living with the virus. As a touristic and commercial hub in Latin America, Brazil harbors an elevated HIV genetic variability, further contributed by the selective pressure exerted by the host immune system and by antiretroviral treatment. Through the progress of the next-generation sequencing (NGS) techniques, it has been possible to expand the study of HIV genetic diversity, evolutionary, and epidemic processes, allowing the generation of HIV complete or near full-length genomes (NFLG) and improving the characterization of intra- and interhost diversity of viral populations. Greater sensitivity in the detection of viral recombinant forms represents one of the major improvements associated with this development. It is possible to identify unique or circulating recombinant forms using the near full-length viral genomes with increasing accuracy. It also permits the characterization of multiple viral infections within individual hosts. Previous Brazilian studies using NGS to analyze HIV diversity were able to identify several distinct unique and circulating recombinant forms and evidenced dual infections. These data unveiled unprecedented high rates of viral recombination and highlighted that novel recombinants are continually arising in the Brazilian epidemic. In the pooled analysis depicted in this report, HIV subtypes have been determined from HIV-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul). Combined data analysis showed a significant prevalence of recombinant forms (29%; 101/350), and a similar 26% when only NFLGs were considered. Moreover, the analysis was able to evidence the occurrence of multiple infections in some individuals. Our data highlight the great HIV genetic diversity found in Brazil and unveils a more accurate scenario of the HIV evolutionary dynamics in the region.

**Keywords:** HIV-1, genetic diversity, NGS, NFLG, subtype

## INTRODUCTION

The first registry of AIDS was reported in the beginning of the 1980s, and until now 77 million people have become infected with human immunodeficiency virus (HIV) and 35 million have died from AIDS-related causes. By 2017, it was estimated that 36.9 million people worldwide were living with HIV (UNAIDS, 2018). Harboring over one-third of the total population of Latin America, Brazil accounts for nearly half of the new HIV infections and of the estimated total of individuals living with HIV/AIDS (48 and 46%, respectively) in the region (UNAIDS, 2018). This scenario, along with the high error-prone rate of the viral reverse transcriptase (RT), high virus replication rates and recombination events, contributes to the remarkable accumulation of genetic diversity in its population during the course of infection, further influenced by selective pressure exerted by the host immune system and by antiretroviral treatment (Roberts et al., 1988; Overbaugh and Bangham, 2001; Zhuang et al., 2002; Santoro and Perno, 2013).

The surveillance of HIV diversity assists to monitor the emergence of new subtypes and the presence of novel strains in a given geographic location (Hemelaar, 2013). The great diversity of HIV-1 group M, which disseminated on a global scale and dominates the current AIDS pandemic, allowed the phylogenetic classification in nine pure subtypes (A–D, F–H, J, and K), sub-subtypes (A1–A5, F1–F2), circulating recombinant forms (CRFs) and unique recombinant forms (URFs) (Robertson et al., 2000). Currently, more than 90 CRFs have been reported in the HIV Sequence Database of the United States Los Alamos National Laboratory<sup>1</sup>. Recombinant viruses are the result of simultaneous infection by multiple viruses during a single transmission event (co-infection) or from sequential infection at multiple transmission events (superinfection) (Yerly et al., 2004). Molecular epidemiology studies show that the overall distribution of HIV-1 groups, subtypes and recombinant forms is highly heterogeneous, with significant differences in the size of the epidemic and the geographical distribution. Overall, subtype C is responsible for half of the current infections (48%), followed by subtypes A (12%), and B (11%). A high prevalence of recombinant forms, which account for at least 21% of HIV-1 infections worldwide, is also noteworthy (Hemelaar, 2012). As seen in Latin America and the Caribbean countries, subtype B prevails in most parts of Brazil, followed by subtypes F1, C, D, and diverse recombinant forms. Southern Brazil, however, presents a distinct epidemiological pattern, with a higher prevalence of subtypes C, B, and BC recombinants (Cardoso et al., 2009; Machado et al., 2009, 2017; de Medeiros et al., 2011; Almeida et al., 2012; Graf and Pinto, 2013; Velasco-de-Castro et al., 2014; Junqueira and Almeida, 2016; Pessoa et al., 2016; Delatorre et al., 2017; Filho and Brites, 2017; Lima et al., 2017).

Through the progress of next-generation sequencing (NGS) techniques it became possible to expand the study of HIV genetic diversity, evolutionary and epidemic processes, allowing the generation of HIV complete or near full-length genomes (NFLG) and improving the characterization of intra- and interhost

diversity of viral populations. Greater sensitivity and accuracy in the detection of viral recombinant forms represents one of the major improvements associated with this development, since most of the previous studies were based on partial HIV genomic sequences, mainly within the *pol* gene region due to the interest in determining drug resistance mutational patterns, resulting in underestimation of the frequency of recombinant forms (Thomson and Najera, 2005; Hemelaar et al., 2011; Marques et al., 2018). It also permits the characterization of multiple viral infections within individual hosts. Previous Brazilian studies using NGS to analyze HIV diversity were able to identify several distinct unique and circulating recombinant forms and evidenced dual infections (Pessoa et al., 2014b, 2015, 2016; Alves et al., 2017; Marques et al., 2018). In the pooled analysis depicted in this report, we pooled publically available Brazilian sequences obtained by NGS and new genetic data from HIV-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul). Upon analyzing the HIV subtypes of this large cohort, our study was able to unveil, with unprecedented accuracy, high rates of viral recombination and highlighted that novel recombinants are continually arising in the Brazilian epidemic.

## MATERIALS AND METHODS

### Study Population and Sample Collection

A total of 84 convenience samples were used for generating the experimental data presented in this study. These were from HIV-1-seropositive patients recruited between February 2016 and December 2017 during the routine services conducted at Sexually Transmitted Diseases/HIV ambulatory at Hospital Federal de Ipanema (HFI) and at Hospital Universitário Clementino Fraga Filho (HU-UFRJ), both located in Rio de Janeiro, southeastern Brazil, and at Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU-FURG), located in Rio Grande, southern Brazil. Clinical and epidemiological data were obtained through a questionnaire and 10 ml of whole peripheral blood were collected. This research was approved by the Ethics Committees in Research of the Brazilian National Cancer Institute – INCA and of HFI (CAAE 52862016.9.0000.5274), HUCFF-UFRJ (CAAE 56604816.2.0000.5257), and HU-FURG (CAAE 52862016.9.3001.5324). The inclusion criteria were being 18 years or greater, being under first antiretroviral scheme and being upon virological success (undetectable HIV viral load) for the last 12 months. A fraction of this casuistic, 32 patients from HFI, has been previously described (Alves et al., 2017). We pooled these data with all Brazilian HIV-1 data comprising NFLG and partial sequences determined by NGS and publically available (Pessoa et al., 2015, 2016). All studies included in the present report (either experimentally determined herein or retrieved from the literature) used a similar methodology to amplify the HIV NFLG and sequence them in an Illumina MiSeq platform. Multiple infection analyses were also made by *de novo* assembly as described below.

<sup>1</sup><http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>

## DNA Extraction and PCR of Proviral DNA

The patients' genomic DNA containing their HIV-1 proviruses was extracted from whole blood with the Genomic DNA Extraction Kit (Real Genomics, BioAmerica, Inc.) following manufacturer's instructions. Nested PCR performed using Platinum™ Taq DNA Polymerase High Fidelity (Life Technologies) was carried out in a Veriti® 96-Well Thermal Cycler (Life Technologies, Carlsbad, United States) for the amplification of HIV NFLGs. The strategies comprised the amplification of four to five fragments, from 2 to 3 kb each, spanning the whole HIV genome (Sanabani et al., 2006b; Ode et al., 2015). After visualization using GelRed (Biotium, Hayward, CA, United States) in 1% agarose gels, duplicated independent PCR-positive products were pooled directly to avoid representativeness of PCR-based errors and their interpretation as minority variants in the population. PCR products were purified with the GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, MA, United States) and their concentration was measured in a NanoDrop ND 1000 apparatus (Thermo Fisher Scientific, MA, United States). The purified products were diluted to 4 ng/μL and pooled per patient.

## Library Construction and NGS

Libraries were prepared with the Nextera XT DNA Sample Preparation kit (Illumina Inc., San Diego, United States) according to the manufacturer's protocol, except that the starting material was diluted to 0.4 ng/μL. The library construction consists of a fragmentation step using transposon technology, followed by a PCR step where dual indexes were added to the fragments. After this process, libraries were quantified by qPCR with the KAPA library quantification kit (Kapa Biosystems, MA, United States) or by fluorometric quantitation with the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, EUA). Libraries were diluted, and pooled prior to sequencing in a MiSeq Illumina platform (2 × 301 cycles paired-end run) (Illumina) with 1% denatured PhiX DNA as a sequencing control.

## Data Analysis

The analysis of the obtained files was performed in Geneious v.9.1.3 using the same alignment parameters for a reference-based genome assembly described by Dudley et al. (2014). Reads were assembled using an annotated HIV-1 HXB2 reference sequence and 10 iterations to obtain each viral genome sequence. The presence of multiple infections was investigated by *de novo* assembly performed with IVA (*Iterative Virus Assembler*) with default parameters (Hunt et al., 2015). This assembler is based on seed sequences that are iteratively and conservatively extended into a contig using reads that have a perfect match. After this process, the program uses those contig as reference in a reference genome aligner to extend the initial contigs obtained (Hunt et al., 2015). The contig sequences obtained were submitted to a BLAST nucleotide analysis to discard those of human origin and confirmed with the Los Alamos National Laboratories HIV BLAST Tool<sup>2</sup>. All samples suggestive of multiple infections (more than one contig representing the same genomic region)

were re-aligned using their respective IVA-assembled contigs as references in the Geneious program. More restrictive assembly parameters were used, such as allowing mapping of paired reads only when both mapped to the same contig and discarding reads that mapped simultaneously to more than one contig. To investigate if these contigs were derived from the same virus, sequences were submitted to hypermutation analysis using the Hypermut2.0 tool available at Los Alamos HIV Database. The contigs were considered hypermutated if the *p*-value was  $\leq 0.05$  when comparing the number of APOBEC G-to-A signature mutations with the control context. The overlapped region between the contigs obtained per sample was compared to all sequences publically available at the BLASTn Database<sup>3</sup>. The top 10 hits of each contig were retrieved from this database to construct phylogenetic trees and check its clustering profile.

## Phylogenetic Analysis

The consensus sequence for each sample was extracted from the reference-guided assembly described above using the 50% stringency setting and classified using maximum likelihood phylogenetic analysis performed with PhyML v.3.0 and the best model of nucleotide substitution defined with Model Generator (Keane et al., 2006; Guindon et al., 2010). To investigate HIV-1 recombination the *bootscanning* tool of Simplot v.3.5.1 was used with the following parameters: *window* = 400 pb; *steps* = 40 pb; *T/t* = 2.0; *gapstrip* = on; *replicas* = 100; *nucleotide substitution model* = F84; *method* = Maximum Likelihood (Lole et al., 1999). Phylogenetic analyses were repeated for recombinant sequences considering the *bootscanning* breakpoint analysis (data not shown). The sequences obtained in this study were submitted to the GenBank under the accession numbers MK041550-MK041589. The raw sequencing reads were deposited to the Sequence Read Archive (SRA) under the numbers SRR7993842-SRR7993872.

## RESULTS

In this study, we included data previously published by our group from patients followed-up at HFI and all HIV-1 sequences obtained by NGS publicly available. A detailed description of the studied populations can be found in their respective articles (Pessoa et al., 2015, 2016; Alves et al., 2017). We focused on the HIV-1 subtype classification, identification of HIV recombinants and multiple infection investigation for our novel patients. As previous published, the patients from HFI were mostly males (75%) with a median age of 38 years at the time of sample collection. Regarding our new cohort, we also found a prevalence of males (67%) among the patients from HU-UFRJ with a median age of 43.5 years. Unlike the other centers, a greater number of female patients (62%) were observed among patients from HU-FURG, with a median age of 43 years. Clinical and epidemiological characteristics of the three cohorts are compiled in **Table 1**. Regarding antiretroviral treatment, 19 patients from HI (19/32, 59%),

<sup>2</sup>[https://www.hiv.lanl.gov/content/sequence/BASIC\\_BLAST/basic\\_blast.html](https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html)

<sup>3</sup><https://blast.ncbi.nlm.nih.gov/>

**TABLE 1** | Demographic and clinical characteristics of HIV-positive patients.

Characteristic	HI (n = 32)	HU- UFRJ (n = 12)	HU-FURG (n = 40)
Males (%)	24 (75%)	8 (67%)	15 (38%)
Age (years)	38	43.5	43
(median; IQR <sub>50</sub> )	(31.5–45.25)	(41–52)	(35–49.5)
Median baseline CD4+ T-cell counts (cells/mm <sup>3</sup> ; IQR <sub>50</sub> )	712.5 (606.5–856)	1117 <sup>1</sup>	780.5 (680.75–938.5)
Median baseline CD8+ T-cell counts (cells/mm <sup>3</sup> ; IQR <sub>50</sub> )	657.5 (529–1047.25)	NA	848 (750.5–1053)
Median time since HIV diagnosis (years; IQR <sub>50</sub> )	4.7 (3.9–6.5)	13.3 (7.7–14.8)	4.83 (2.42–10.9)
Median time from HIV diagnosis to start of ARV treatment (years; IQR <sub>50</sub> )	1.2 (0.6–2.8) <sup>2</sup>	1.7 (0.1–5.3)	0.88 (0.25–3.2)
Median treatment time (years; IQR <sub>50</sub> )	3.1 (2.4–3.9)	9.34 (4.6–11.4)	3.83 (2.2–4.77)

<sup>1</sup>Information available for only two patients (1201 and 1033 cells/mm<sup>3</sup>). <sup>2</sup>Data based on 31 patients. NA, information not available. IQR<sub>50</sub>, interquartile range.

four patients from HU-UFRJ (4/12, 33%), and 23 patients from HU-FURG (23/40, 58%), were under the HAART scheme preconized by Brazilian Ministry of Healthy at the time of sample collection, composed of tenofovir (TDF), lamivudine (3TC), and efavirenz (EFV). All schemes used by the patients are described in **Table 2**.

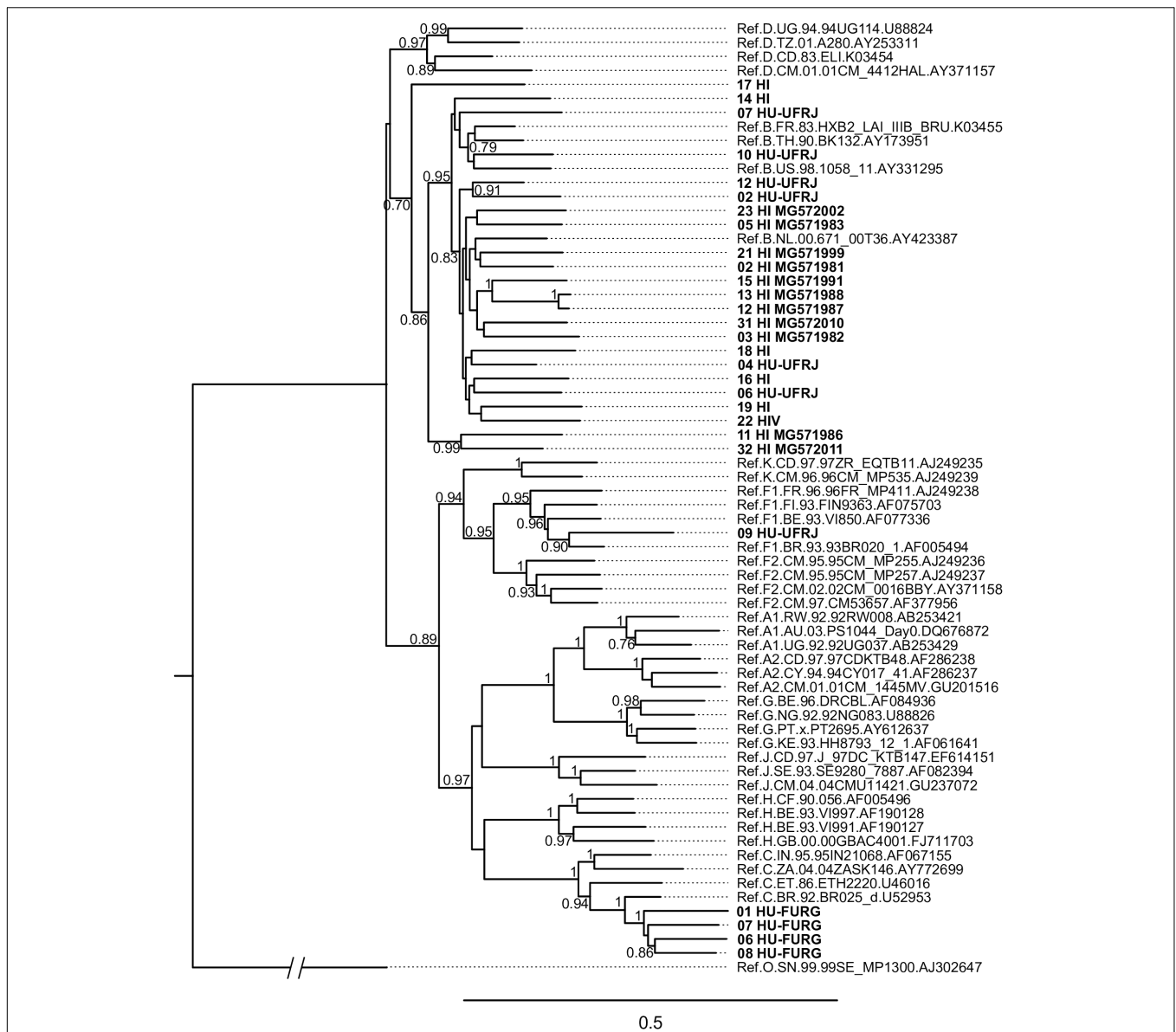
Overall, we were successful in sequencing by NGS at least 2 of the 4/5 overlapping PCR fragments of 48 samples from our cohort included in this study (57%, 48/84). Of those, 28 samples (58%; 28/48) had the NFLG obtained. The remaining partial genome sequences had complete Gag CDS (coding sequence) for nine samples (45%, 9/20), Pol CDS for 4 (20%, 4/20), and

**TABLE 2** | Distribution of subtypes and HAART regimen exposure across the 48 HIV-1 genome sequences analyzed.

Patient	Subtype/ URF	HAART Regimen	Patient	Subtype/ URF	HAART Regimen
1-HI	B	AZT+3TC+NVP	<b>31-HI</b>	<b>B</b>	TDF+3TC+EFV
<b>2-HI#</b>	<b>B</b>	AZT+3TC+EFV	<b>32-HI</b>	<b>BF</b>	TDF+3TC+EFV
<b>3-HI</b>	<b>B</b>	TDF+3TC+EFV	<b>2-HU-UFRJ</b>	<b>B</b>	TDF+3TC+EFV
4-HI	B	TDF+3TC+EFV	3-HU-UFRJ	B	AZT+3TC+EFV
<b>5-HI</b>	<b>B</b>	AZT+3TC+ATV	<b>4-HU-UFRJ</b>	<b>B</b>	AZT+3TC+EFV
6-HI	F1	AZT+3TC+ATV/r	<b>6-HU-UFRJ</b>	<b>B</b>	AZT+3TC+EFV
8-HI	BC	TDF+3TC+EFV	<b>7-HU-UFRJ</b>	<b>B</b>	AZT+3TC+EFV
<b>11-HI</b>	<b>BF</b>	TDF+3TC+EFV	<b>9-HU-UFRJ</b>	<b>F1</b>	AZT+3TC+AZT/r
<b>12-HI</b>	<b>B</b>	TDF+3TC+EFV	<b>10-HU-UFRJ</b>	<b>B</b>	TDF+3TC+EFV
<b>13-HI</b>	<b>B</b>	TDF+3TC+EFV	<b>12-HU-UFRJ</b>	<b>B</b>	TDF+3TC+EFV
<b>14-HI</b>	<b>B</b>	AZT+3TC+LPV/r	<b>1-HU-FURG</b>	<b>C</b>	TDF+3TC+EFV
<b>15-HI</b>	<b>B</b>	TDF+3TC+EFV	2-HU-FURG	C	TDF+3TC+EFV
<b>16-HI</b>	<b>B</b>	AZT+3TC+FPV/r	3-HU-FURG	BC	AZT+3TC+EFV
<b>17-HI</b>	<b>BF1</b>	AZT+3TC+ATV/r	<b>6-HU-FURG</b>	<b>C</b>	TDF+3TC+AZT/r
<b>18-HI</b>	<b>B</b>	TDF+3TC+EFV	<b>7-HU-FURG</b>	<b>C</b>	TDF+3TC+DRV/r
<b>19-HI</b>	<b>B</b>	AZT+3TC+EFV	<b>8-HU-FURG</b>	<b>C</b>	TDF+3TC+ATV
20-HI	B	TDF+3TC+EFV	10-HU-FURG	C	TDF+3TC+EFV
<b>21-HI</b>	<b>B</b>	TDF+EFV+FTC	12-HU-FURG	C	TDF+3TC+EFV
<b>22-HI</b>	<b>B</b>	TDF+3TC+EFV	13-HU-FURG	C	TDF+3TC+EFV
<b>23-HI</b>	<b>B</b>	TDF+3TC+EFV	14-HU-FURG	BF1	AZT+3TC+AZT/r
26-HI	B	TDF+3TC+EFV	16-HU-FURG	C	TDF+3TC+EFV
27-HI	B	TDF+3TC+EFV	17-HU-FURG	C	AZT+3TC+LPV/r
28-HI	BF	TDF+3TC+EFV	18-HU-FURG	BF1	TDF+3TC+EFV
29-HI	B	TDF+3TC+EFV	20-HU-FURG	BC	TDF+3TC+EFV

#The near full-length genomes are represented in bold.



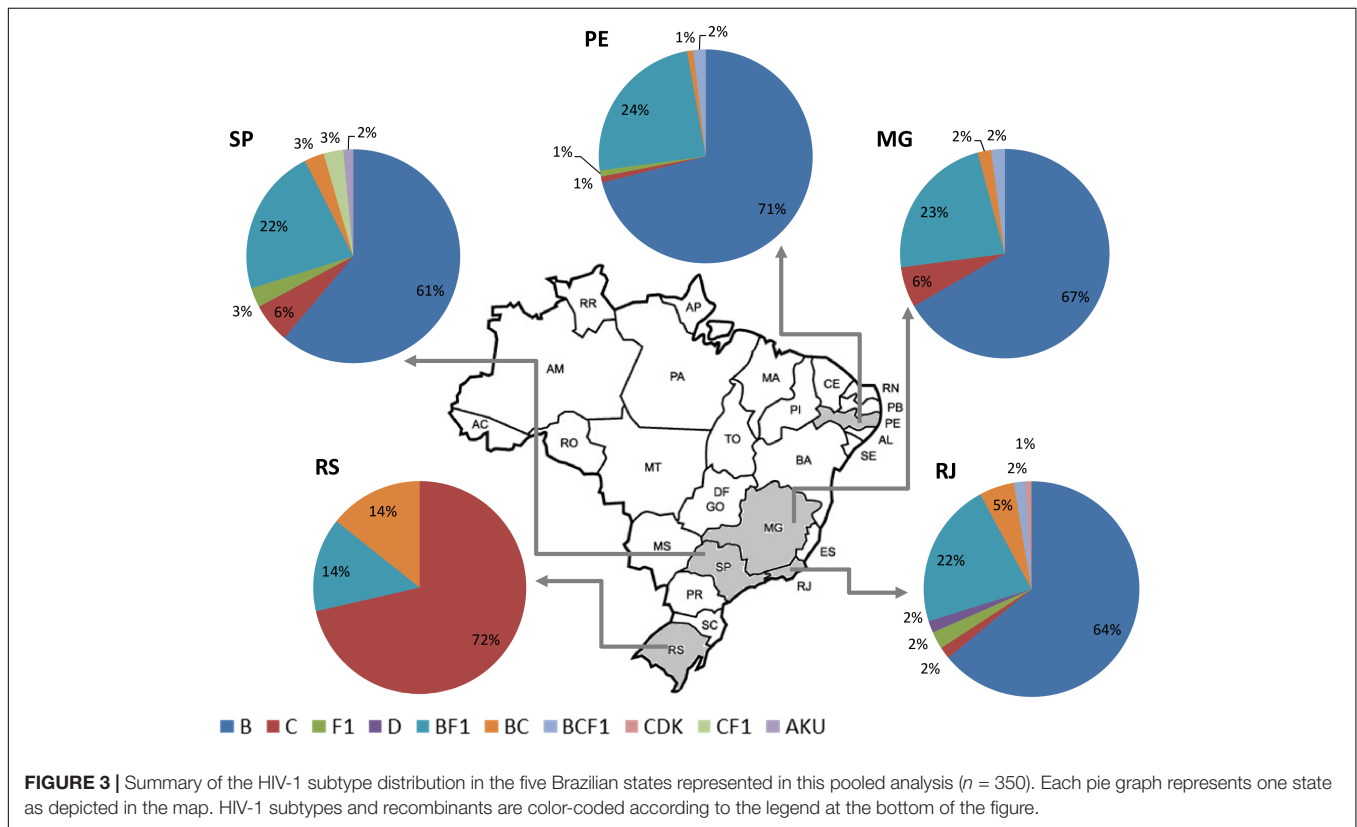
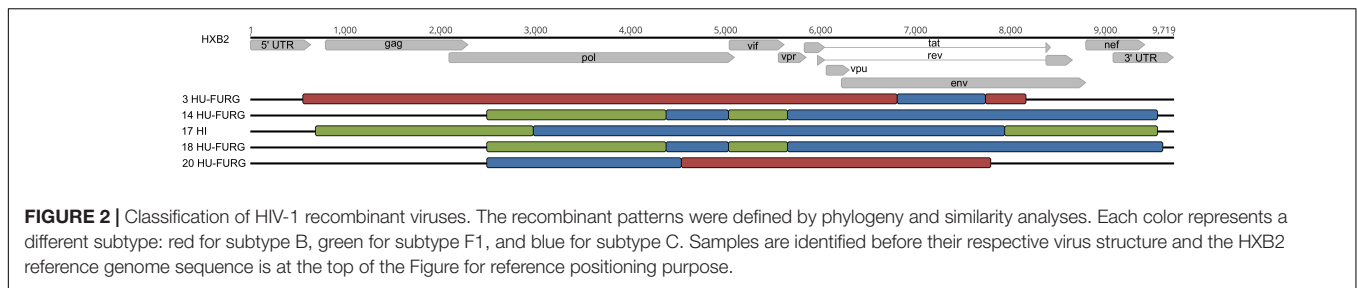


**FIGURE 1 |** Phylogenetic analysis of HIV near full-length genomes (NFLG) from this study. The maximum likelihood analysis was performed with 1,000 bootstrap iterations. The tree contains 28 HIV-1 proviral sequences obtained from Hospital Federal de Ipanema (HFI), Hospital Universitário Clementino Fraga Filho-UFRJ (HU-UFRJ), Rio de Janeiro, and Hospital Universitário Dr. Miguel Riet Corrêa Jr. (HU-FURG), Rio Grande do Sul (represented in bold) and reference sequences of HIV-1 subtypes (named by subtype, country, year, and GenBank accession number). HI sequences determined in a previous study (Alves et al., 2017) are named with their respective GenBank accession numbers. Only bootstrap values greater than 0.7 are shown.

Env CDS for 12 (60%, 12/20). Six samples from HI, 4 from HU-UFRJ, and 26 from HU-FURG (43%) failed to have more than one viral DNA fragment PCR-amplified and were excluded from further analyses. Of the 302 NGS Brazilian sequences previously available in the literature, which information was also included in this study, 247 were NFLG and 55 were partial sequences (Pessoa et al., 2015, 2016).

HIV-1 consensus sequences were subjected to phylogenetic analysis to determine their subtype/CRF classification. The NFLGs obtained by our group were mostly classified as HIV-1 subtype B (71%; 20/28), followed by subtype C (14%; 4/28),

recombinant forms (11%; 3/28), and subtype F1 (4%; 1/28) (Figure 1). Table 2 describes the subtype classification of each sample included. The recombinants were classified as distinct URFs involving subtypes B and F1 based on Simplot analysis (Figure 2). Two of them, HI-11 and HI-32, were already described in our former study (Alves et al., 2017). The sequences of the 20 HIV-1 partial genomes were predominantly subtype B (7/20), C (6/20), and recombinant forms (6/20) (Table 2). A single subtype F1 sample was identified. With respect to the recombinants found in our study and submitted to Simplot analysis, two were identical URF\_BF1 and a third one was a

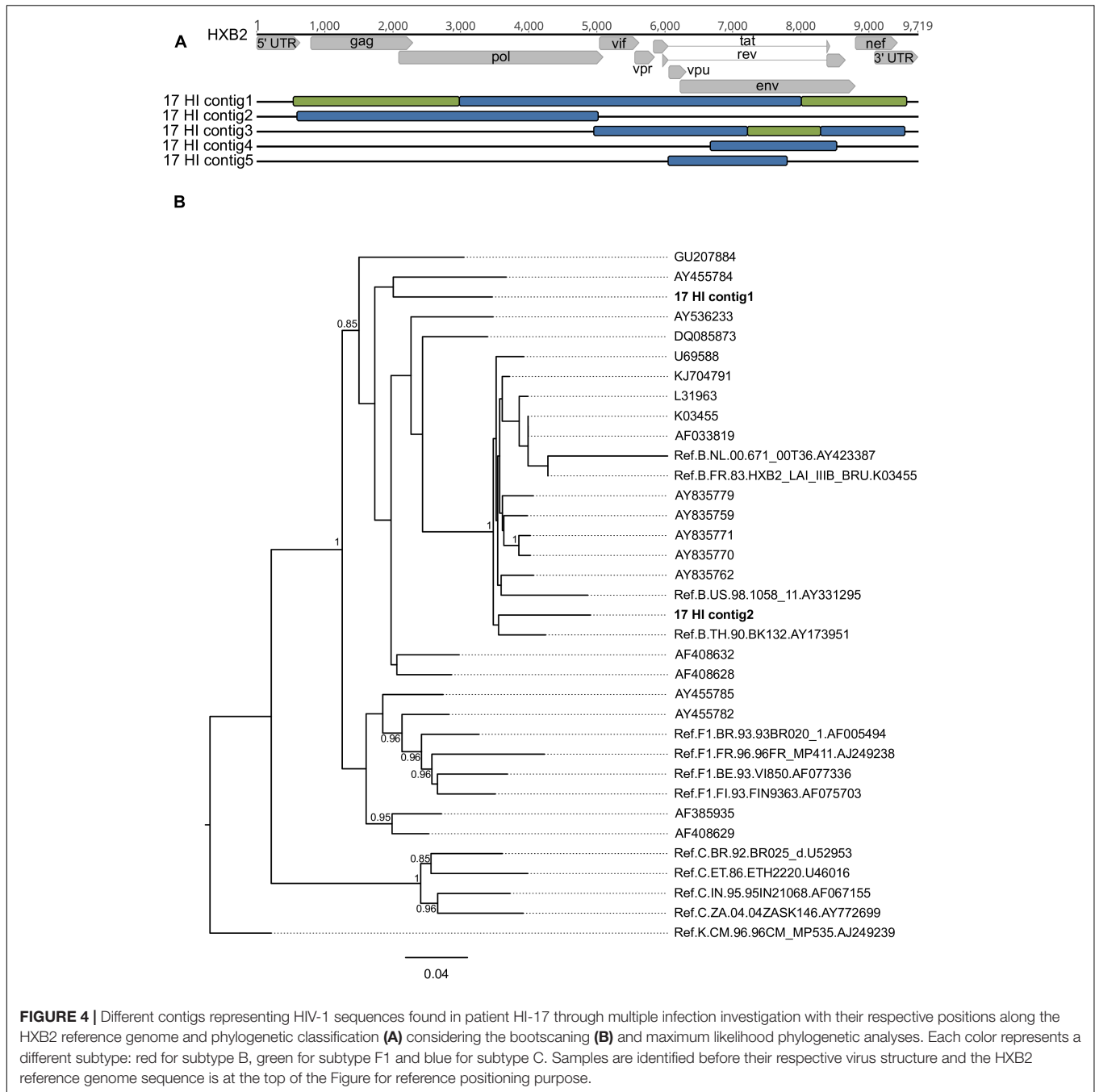


unique URF\_BF1 previously described (HI-28), and three were distinct URF\_BC strains (one of those also described, HI-08) (Figure 2; Alves et al., 2017).

Altogether, the pooled analysis of all Brazilian NGS data, including NFLG (29%, 79/275) and partial sequences (29%, 22/75), showed a significant prevalence of recombinant forms (29%; 101/350). Considering the distribution of the HIV-1 subtypes and recombinant viruses in the five Brazilian states analyzed, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul), we could observe a highly diversified pattern of HIV-1 subtype distribution (Figure 3). A higher prevalence of recombinant forms in São Paulo and Rio de Janeiro (30%) could also be found, followed by Rio Grande do Sul (28%), and by Minas Gerais and Pernambuco (27%), although those differences were not significant (data not shown).

*De novo* analysis generated results suggestive of multiple infections (more than one IVA-contig in the same genomic

region) for four samples (HU-UFRJ-03, two contigs; HI-11, two contigs; HI-14, three contigs, and HI-17, five contigs). For three of them (HU-UFRJ-03, HI-11, and HI-14), only one of the contigs generated intact open reading frames (ORFs), while the other contigs presented truncated ORFs showing multiple stop codons, consistent with APOBEC-mediated G-to-A mutations. They were confirmed as hypermutated sequences when compared to the viable sequence from the respective patient in Hypermut ( $p < 0.05$ , data not shown). Patient HI-17 had several overlapping regions between the contigs, two of them at the *gag-pol* region and three at *env* (Figure 4A). Phylogenetic trees comprising the contigs, the top-ten best hits found in BLASTn searches for each contig and reference sequences were constructed for each overlapping region. Overall, these trees showed different clustering profiles between the contigs and HIV-1 subtype references, suggesting the presence of variants with different HIV-1 subtypes within this samples (Figure 4B). The contigs were then submitted to Simplot



analysis to confirm the subtype classification and determine the recombination breakpoint profile. All contigs were confirmed as distinct variants. For the HI-17 patient, the longest contig had the same recombination profile observed for the consensus sequence (URF\_BF1), one had a distinct recombination profile comprising B and F1 subtypes, and three were classified as subtype B (Figure 4A). Another two cases of multiple infections with distinct subtypes were described by Pessoa et al., one involving viruses of subclade F1 and subtype B and another involving a CBF1 recombinant and a non-recombinant subtype B (Pessoa et al., 2015, 2016).

## DISCUSSION

The present study describes the HIV-1 genetic diversity and molecular epidemiology observed in Brazil using NGS-generated HIV-1 sequences, combining recently published reports and novel data from our group. In comparison to our previous published study with HIV+ patients recruited at HFI, we extended our analysis for three of the nine patients which data could not be obtained in the first study, and obtained NFLG sequences for another five patients (Alves et al., 2017). Concerning the integrity of the open reading frames (ORF),

4.5% (11/247) of the sequences available in the literature, and 25% (7/28) of our sequences displayed mutations and/or insertions and deletions resulting in frameshifts or premature stop codons.

Considering our cohort, a high prevalence of infection by HIV-1 subtype B viruses was found among patients from Rio de Janeiro (79%, 27/34). On the other hand, subtype C was the most prevalent subtype in Rio Grande do Sul, with a similarly high proportion (72%, 10/14), highlighting the regional differences observed in HIV-1 subtype distribution in the country. It is well documented that the overall prevalence of non-B strains, such as URF\_BF1, URF\_BC, and particularly subtype C and CRF31\_BC in the South of Brazil, has been increasing (Santos et al., 2007; Cardoso et al., 2009; Machado et al., 2009; Almeida et al., 2012; Velasco-de-Castro et al., 2014). Similarly to some recently published data using NFLG, the recombinants identified in our cohort did not show any similarity with the CRFs already described (Sanabani et al., 2013; Alves et al., 2017). However, it is worth to mention that two partial sequences from HU-FURG show the same recombinant pattern (14 HU-FURG and 18 HU-FURG) and did not have any evidence of epidemiological linkage between them.

Through the analysis of HIV-1-positive patients in five states of Brazil with some of the highest HIV prevalence, three in the Southeast (Rio de Janeiro, São Paulo, and Minas Gerais), one in the Northeast (Pernambuco) and one in the South (Rio Grande do Sul), we could determine, with unprecedented accuracy, the prevalence of HIV-1 recombinant forms in the Brazilian epidemic and their distribution across the country. The high prevalence of recombinant strains (29%) identified by NGS is supported by the circulation of multiple subtypes and consistent with the hypothesis that novel recombinants are continuously arising in the Brazilian epidemic (Sanabani et al., 2013). Our cohort presented 19% of recombinant sequences, a lower prevalence than 28 and 40% observed in other Brazilian NGS-based studies (Pessoa et al., 2015, 2016). The higher prevalence of recombinant forms when compared to recent Brazilian Sanger sequencing-based studies (ranging from 5 to 16%) can be attributed to the smaller genomic region analyzed in the latter, mostly based only on *pol* gene, which impairs the accurate classification of recombinants (Librelotto et al., 2015; Moura et al., 2015; da Costa et al., 2016; Dos Anjos Silva et al., 2016; Delatorre et al., 2017; Filho and Brites, 2017; Lima et al., 2017). Some of these studies covering small genomic regions could not find recombinant strains, like the study conducted by Corado et al. (2017) among 73 individuals from Roraima state, northern Brazil. A study conducted by Graf et al. (2016) found a higher proportion of recombinant strains than others Sanger-based studies (21%, 66/317). However, this study used molecular data from more than one gene (HIV-1 *pol*, *env*, or both). It should also be noted that 30% of the URF\_BC samples were intergenic recombinants whose recombination breakpoints were not documented within these fragments (Graf et al., 2016). The comparison of the estimated prevalence of recombinant virus between the classical Sanger-based approach and the NGS data clearly highlights underestimated rates in the former analyses, mainly associated with the smaller

genomic region analyzed, which also implies in the inaccurate detection of the recombinant breakpoints and, therefore, their classification.

HIV-1 NFLG-based studies can unveil an underestimated rate of recombinant viruses in the country. Using NFLG data, several studies have described new Brazilian CRF strains designated as CRF28\_BF, CRF29\_BF, CRF39\_BF, CRF40\_BF, CRF46\_BF, and CRF31\_BC, CRF70\_BF1, CRF71\_BF1, CRF72\_BF1, CRF90\_BF1 (De Sa Filho et al., 2006; Sanabani et al., 2006a,b, 2010; Santos et al., 2006; Guimaraes et al., 2008; Pessoa et al., 2014a,b; Reis et al., 2017). An important study conducted by Pessoa et al. evaluated the complete genomes of HIV-1 strains by NGS previously sequenced by Alencar et al. assigned to subtype F1 and showed that 23 of the 24 samples analyzed were BF recombinants, with 4 CRF70\_BF1 and 11 CRF71\_BF1 novel recombinant types (Alencar et al., 2013; Pessoa et al., 2014b). The same was observed by Marques et al. (2018) where 34 of the 55 sequences analyzed were classified as recombinants. In our cohort, only one sequence was classified as subtype F1 and three as URF-BF1, which corroborates to literature and highlights the higher prevalence of recombinants forms comprising subtype F1.

Regarding molecular diversity, our pooled analysis corroborates the crescent prevalence of non-B strains in the Brazilian epidemic, confirming the phylogenetic intermixing of HIV-1 sequences. The most prominent case comprises subtype C and C-containing recombinant forms expanding from the South of Brazil to other regions (Bello et al., 2012; Graf and Pinto, 2013; Graf et al., 2015). Non-B strains represent 39% of the sequences from São Paulo, 36% from Rio de Janeiro, 33% from Minas Gerais, and 29% from Pernambuco. Subtype B was not found in Rio Grande do Sul, probably because of the small number of samples analyzed in this region.

We also investigated infection by distinct variants using a *de novo* strategy to obtain sequences from each patient that are subsequently run in one of the reference-guided approaches using this sequence as a reference (see section Data Analysis of Materials and Methods). This strategy was employed by several studies to reduce the influence of a reference genome in the assembly process while investigating multiple infections (Mangul et al., 2014; Aralaguppe et al., 2016; Alampalli et al., 2017; Baaijens et al., 2017). The prevalence of multiple infections observed in our study (2%, 1/48) was similar to the prevalence reported by Pessoa et al. (4%, 1/24 and 2%, 1/47), but both were greater than the prevalence subsequently reported (0.3%, 1/259) (Pessoa et al., 2014b, 2015, 2016, respectively). At this point, it is not possible to infer whether the distinct viral strains resulted from coinfections or the acquisition of a second variant after the establishment of the first one (superinfections).

We are aware that the pooled analysis presented here includes HIV-positive patients with different HIV clinical profiles. While our cohort is composed by patients under first-line HAART and undetectable HIV viral load for at least 12 months prior to collection date attending at sexually transmitted diseases/HIV ambulatory, the studies conducted by Pessoa et al. involved recently infected donors at four blood centers. However, it should be noted that all epidemiological HIV-1 NFLG studies based on NGS conducted in Brazil available in the literature were



included in this pooled analysis. However, it should be noted that our convenience samples may have biased the analysis, especially regarding the multiple infection prevalence, which requires analysis from a larger data set.

Like previously described, the inclusion criteria used in this study were very strict and represented an important barrier to enroll a large number of patients. The difficulty at PCR-amplification of archived proviral genomes was also a limitation due to early chronic infection and undetectable HIV viral load. It is also important to mention that HU-FURG samples were incorporated in this study at a later stage. This fact, coupled with the difficulty in the PCR amplification of these samples resulted in a limited number of sequences from Rio Grande do Sul and additional studies are necessary to complement our findings. Another limitation of our study was that only the NFLG sequences ( $n = 28$ ) were evaluated for ORF intactness, and seven of them (25%) had stop codons due to hypermutation or to frameshift deletions.

The analysis of all Brazilian HIV-1 NFLG obtained by NGS give us a more accurate evaluation of the viral diversity present in this epidemic. Through the subtype analyses conducted in this large cohort, we were able to find high rates of viral recombination, showing that larger viral genomic regions are required for reliable genetic evaluation and thus to establish effective public health policies to assure suitable HIV screening, diagnosis, monitoring and novel strategies based on viral variability. Our data highlight the great HIV genetic diversity found in Brazil and unveils a more accurate scenario of the HIV evolutionary dynamics in the region.

## ETHICS STATEMENT

This study was carried out and approved in accordance with the recommendations of the Ethics Committees in Research of the Brazilian National Cancer Institute – INCA and of HFI (CAAE 52862016.9.0000.5274), HUCFF-UFRJ (CAAE 56604816.2.0000.5257), and HU-FURG (CAAE

52862016.9.3001.5324), with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

BA, ES, and MS conceived the study. BA and VDH collected and processed the samples. BA, JS, IP, and OB performed all molecular biology experiments and all the bioinformatics analyses. SS contributed with previously published HIV-1 NGS sequences from Brazil. PR-P and HF contributed reagents and provided critical reading of the manuscript. BA, JS, and MS wrote the manuscript. All authors read the manuscript and agreed with its final version and submission.

## FUNDING

This work was supported by the Rio de Janeiro State Science Foundation – FAPERJ (Grant # E-26/201.331/2014 to MS), by the Brazilian National Research Council – CNPq (Grants # 454623/2014-3 to MS and # 460346/2014-8 to ES), and by MSD Avenir (Grant DS-2016-0005 to HF with a subcontract to MS). Intramural funding by the Brazilian Ministry of Health through INCA in the form of fellowships was provided to BA and JS at the post-doc level, to IP at the Ph.D. level and to OB at the undergraduate level.

## ACKNOWLEDGMENTS

We would like to thank the medical staff personnel of Hospital Federal de Ipanema (Marianne M. Garrido and Sayonara R. Ribeiro), Hospital Universitário Dr. Miguel Riet Corrêa Jr. (Jussara Silveira and Rossana Basso), and Hospital Universitário Clementino Fraga Filho (Elizabeth S. Machado) for providing clinical follow-up and information on the studied patients at their respective clinical sites.

## REFERENCES

- Alampalli, S. V., Thomson, M. M., Sampathkumar, R., Sivaraman, K., Aj, U. K. J., Dhar, C., et al. (2017). Deep sequencing of near full-length HIV-1 genomes from plasma identifies circulating subtype C and infrequent occurrence of AC recombinant form in Southern India. *PLoS One* 12:e0188603. doi: 10.1371/journal.pone.0188603
- Alencar, C. S., Sabino, E. C., Carvalho, S. M., Leao, S. C., Carneiro-Proietti, A. B., Capuani, L., 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. doi: 10.1097/QAI.0b013e31828ff979
- Almeida, S. E., De Medeiros, R. M., Junqueira, D. M., Graf, T., Passaes, C. P., Bello, G., et al. (2012). Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. *Viol. J.* 9:306. doi: 10.1186/1743-422X-9-306
- Alves, B. M., Siqueira, J. D., Garrido, M. M., Botelho, O. M., Prellwitz, I. M., Ribeiro, S. R., et al. (2017). Characterization of HIV-1 near full-length proviral genome quasispecies from patients with undetectable viral load undergoing first-line haart therapy. *Viruses* 9:E392.
- Aralaguppe, S. G., Siddik, A. B., Manickam, A., Ambikan, A. T., Kumar, M. M., Fernandes, S. J., et al. (2016). Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. *J. Virol. Methods* 236, 98–104. doi: 10.1016/j.jviromet.2016.07.010
- Baaijens, J. A., Aabidine, A. Z. E., Rivals, E., and Schonhuth, A. (2017). De novo assembly of viral quasispecies using overlap graphs. *Genome Res.* 27, 835–848. doi: 10.1101/gr.215038.116
- Bello, G., Zanotto, P. M., Iamarino, A., Graf, T., Pinto, A. R., Couto-Fernandez, J. C., et al. (2012). Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. *PLoS One* 7:e35649. doi: 10.1371/journal.pone.0035649
- Cardoso, L. P., Queiroz, B. B., and Stefani, M. M. (2009). HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naive patients from Central West Brazil. *J. Clin. Virol.* 46, 134–139. doi: 10.1016/j.jcv.2009.07.009
- Corado, A. L., Bello, G., Leao, R. A., Granja, F., and Naveca, F. G. (2017). HIV-1 genetic diversity and antiretroviral drug resistance among individuals from

- Roraima state, northern Brazil. *PLoS One* 12:e0173894. doi: 10.1371/journal.pone.0173894
- da Costa, C. M., Costa De Oliveira, C. M., Chehuan, De Melo, Y. F., Delatorre, E., Bello, G., et al. (2016). High HIV-1 genetic diversity in patients from northern Brazil. *AIDS Res. Hum. Retrovir.* 32, 918–922. doi: 10.1089/AID.2016.0044
- de Medeiros, R. M., Junqueira, D. M., Matte, M. C., Barcellos, N. T., Chies, J. A., and Matos Almeida, S. E. (2011). Co-circulation HIV-1 subtypes B, C, and CRF31\_BC in a drug-naive population from Southernmost Brazil: analysis of primary resistance mutations. *J. Med. Virol.* 83, 1682–1688. doi: 10.1002/jmv.22188
- De Sa Filho, D. J., Sucupira, M. C., Caseiro, M. M., Sabino, E. C., Diaz, R. S., and Janini, L. M. (2006). Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res. Hum. Retroviruses* 22, 1–13.
- Delatorre, E., Couto-Fernandez, J. C., and Bello, G. (2017). HIV-1 Genetic diversity in Northeastern Brazil: high prevalence of non-b subtypes. *AIDS Res. Hum. Retroviruses* 33, 639–647. doi: 10.1089/AID.2017.0045
- Dos Anjos Silva, L., Divino, F., Da Silva Rego, M. O., Lima Lopes, I. G., Nobrega Costa, C. M., et al. (2016). HIV-1 genetic diversity and transmitted drug resistance in antiretroviral treatment-naive individuals from Amapa State, Northern Brazil. *AIDS Res. Hum. Retroviruses* 32, 373–376. doi: 10.1089/AID.2015.0280
- Dudley, D. M., Bailey, A. L., Mehta, S. H., Hughes, A. L., Kirk, G. D., Westergaard, R. P., et al. (2014). Cross-clade simultaneous HIV drug resistance genotyping for reverse transcriptase, protease, and integrase inhibitor mutations by Illumina MiSeq. *Retrovirology* 11:122. doi: 10.1186/s12977-014-0122-8
- Filho, A. W. O., and Brites, C. (2017). Geolocalization of HIV-1 subtypes and resistance mutations of patients failing antiretroviral therapy in Salvador - Brazil. *Braz. J. Infect. Dis.* 21, 234–239. doi: 10.1016/j.bjid.2017.02.006
- Graf, T., Machado Fritsch, H., De Medeiros, R. M., Maletich Junqueira, D., Esteves De Matos Almeida, S., and Pinto, A. R. (2016). Comprehensive characterization of HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS. *J. Virol.* 90, 8160–8168. doi: 10.1128/JVI.00363-16
- Graf, T., and Pinto, A. R. (2013). The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 435, 170–178. doi: 10.1016/j.virol.2012.08.048
- Graf, T., Vrancken, B., Maletich Junqueira, D., De Medeiros, R. M., Suchard, M. A., Lemey, P., et al. (2015). Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *J. Virol.* 89, 12341–12348. doi: 10.1128/JVI.01681-15
- Guimaraes, M. L., Eyer-Silva, W. A., Couto-Fernandez, J. C., and Morgado, M. G. (2008). Identification of two new CRF\_BF in rio de Janeiro State, Brazil. *AIDS* 22, 433–435.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001
- Hemelaar, J. (2013). Implications of HIV diversity for the HIV-1 pandemic. *J. Infect.* 66, 391–400. doi: 10.1016/j.jinf.2012.10.026
- Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and Who-Unaided Network for HIV Isolation, and Characterisation. (2011). Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 25, 679–689. doi: 10.1097/QAD.0b013e328342ff93
- Hunt, M., Gall, A., Ong, S. H., Brenner, J., Ferns, B., Goulder, P., et al. (2015). IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* 31, 2374–2376. doi: 10.1093/bioinformatics/btv120
- Junqueira, D. M., and Almeida, S. E. (2016). HIV-1 subtype B: Traces of a pandemic. *Virology* 495, 173–184. doi: 10.1016/j.virol.2016.05.003
- Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., and McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6:29.
- Librelotto, C. S., Graf, T., Simon, D., De Almeida, S. E., and Lunge, V. R. (2015). HIV-1 epidemiology and circulating subtypes in the countryside of South Brazil. *Rev. Soc. Bras. Med. Trop.* 48, 249–257. doi: 10.1590/0037-8682-0083-2015
- Lima, K., Leal, E., Cavalcanti, A. M. S., Salustiano, D. M., De Medeiros, L. B., Da Silva, S. P., et al. (2017). Increase in human immunodeficiency virus 1 diversity and detection of various subtypes and recombinants in north-eastern Brazil. *J. Med. Microbiol.* 66, 526–535. doi: 10.1099/jmm.0.000447
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., and Novak, N. G. (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.
- Machado, L. F., Costa, I. B., Folha, M. N., Da Luz, A. L., Vallinoto, A. C., Ishak, R., et al. (2017). Lower genetic variability of HIV-1 and antiretroviral drug resistance in pregnant women from the state of Para, Brazil. *BMC Infect. Dis.* 17:270. doi: 10.1186/s12879-017-2392-y
- Machado, L. F., Ishak, M. O., Vallinoto, A. C., Lemos, J. A., Azevedo, V. N., Moreira, M. R., et al. (2009). Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02\_AG in the amazon region of Brazil. *AIDS Res. Hum. Retroviruses* 25, 961–966. doi: 10.1089/aid.2009.0027
- Mangul, S., Wu, N. C., Mancuso, N., Zelikovsky, A., Sun, R., and Eskin, E. (2014). Accurate viral population assembly from ultra-deep sequencing data. *Bioinformatics* 30, i329–i337. doi: 10.1093/bioinformatics/btu295
- Marques, B. C. L., Morgado, M. G., and Guimaraes, M. L. (2018). Potential overestimation of HIV-1 sub-subtype F1 circulation in Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* 113:e170483. doi: 10.1590/0074-02760170483
- Moura, M. E., Da Guarda Reis, M. N., Lima, Y. A., Eulalio, K. D., Cardoso, L. P., and Stefani, M. M. (2015). HIV-1 transmitted drug resistance and genetic diversity among patients from Piauí State, Northeast Brazil. *J. Med. Virol.* 87, 798–806. doi: 10.1002/jmv.24087
- Ode, H., Matsuda, M., Matsuoka, K., Hachiya, A., Hattori, J., Kito, Y., et al. (2015). Quasispecies analyses of the HIV-1 near-full-length genome with illumina MiSeq. *Front. Microbiol.* 6:1258. doi: 10.3389/fmicb.2015.01258
- Overbaugh, J., and Bangham, C. R. (2001). Selection forces and constraints on retroviral sequence variation. *Science* 292, 1106–1109.
- Pessoa, R., Carneiro Proietti, A. B., Busch, M. P., and Sanabani, S. S. (2014a). Identification of a Novel HIV-1 circulating recombinant form (CRF72\_BF1) in deep sequencing data from blood donors in Southeastern Brazil. *Genome Announc.* 2, e386–e314. doi: 10.1128/genomeA.00386-14
- Pessoa, R., Watanabe, J. T., Calabria, P., Felix, A. C., Loureiro, P., Sabino, E. C., et al. (2014b). 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:e112674. doi: 10.1371/journal.pone.0112674
- Pessoa, R., Loureiro, P., Esther Lopes, M., Carneiro-Proietti, A. B., Sabino, E. C., Busch, M. P., et al. (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. doi: 10.1371/journal.pone.0152499
- Pessoa, R., Watanabe, J. T., Calabria, P., Alencar, C. S., Loureiro, P., Lopes, M. E., et al. (2015). Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. *Transfusion* 55, 980–990. doi: 10.1111/trf.12936
- Reis, M., Bello, G., Guimaraes, M. L., and 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:e0178578. doi: 10.1371/journal.pone.0178578
- Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* 242, 1171–1173.
- Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–56.
- Sanabani, S., Kleine Neto, W., Kalmar, E. M., Diaz, R. S., Janini, L. M., and Sabino, E. C. (2006a). 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.
- Sanabani, S., Neto, W. K., De Sa Filho, D. J., Diaz, R. S., Munerato, P., Janini, L. M., et al. (2006b). Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res. Hum. Retroviruses* 22, 171–176.
- Sanabani, S. S., Pastena, E. R., Neto, W. K., Martinez, V. P., and Sabino, E. C. (2010). Characterization and frequency of a newly identified HIV-1 BF1 intersubtype

- circulating recombinant form in Sao Paulo, Brazil. *Virology* 7:74. doi: 10.1186/1743-422X-7-74
- Sanabani, S. S., Pessoa, R., Soares De Oliveira, A. C., Martinez, V. P., Giret, M. T., De Menezes Succi, R. C., et al. (2013). Variability of HIV-1 genomes among children and adolescents from Sao Paulo, Brazil. *PLoS One* 8:e62552. doi: 10.1371/journal.pone.0062552
- Santoro, M. M., and Perno, C. F. (2013). HIV-1 genetic variability and clinical implications. *ISRN Microbiol.* 2013:481314. doi: 10.1155/2013/481314
- Santos, A. F., Schrago, C. G., Martinez, A. M., Mendoza-Sassi, R., Silveira, J., Sousa, T. M., et al. (2007). Epidemiologic and evolutionary trends of HIV-1 CRF31\_BC-related strains in southern Brazil. *J. Acquir. Immune Defic. Syndr.* 45, 328–333.
- Santos, A. F., Sousa, T. M., Soares, E. A., Sanabani, S., Martinez, A. M., Sprinz, E., et al. (2006). Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* 20, 2011–2019.
- Thomson, M. M., and Najera, R. (2005). Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 7, 210–224.
- UNAIDS (2018). *The Joint United Nations Programme on HIV/AIDS (UNAIDS). Data 2018*. Geneva: UNAIDS.
- Velasco-de-Castro, C. A., Grinsztejn, B., Veloso, V. G., Bastos, F. I., Pilotto, J. H., Fernandes, N., et al. (2014). HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. *PLoS One* 9:e87622. doi: 10.1371/journal.pone.0087622
- Yerly, S., Jost, S., Monnat, M., Telenti, A., Cavassini, M., Chave, J. P., et al. (2004). HIV-1 co/super-infection in intravenous drug users. *AIDS* 18, 1413–1421.
- Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., et al. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* 76, 11273–11282.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Alves, Siqueira, Prellwitz, Botelho, Da Hora, Sanabani, Recordon-Pinson, Fleury, Soares and Soares. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.