

# BF Integrase Genes of HIV-1 Circulating in São Paulo, Brazil, with a Recurrent Recombination Region

Atila Iamarino, Fernando Lucas de Melo, Carla Torres Braconi, Paolo Marinho de Andrade Zanotto\*

Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute-ICB II, University of São Paulo, São Paulo, Brazil

## Abstract

Although some studies have shown diversity in HIV integrase (*IN*) genes, none has focused particularly on the gene evolving in epidemics in the context of recombination. The *IN* gene in 157 HIV-1 integrase inhibitor-naïve patients from the São Paulo State, Brazil, were sequenced tallying 128 of subtype B (23 of which were found in non-B genomes), 17 of subtype F (8 of which were found in recombinant genomes), 11 integrases were BF recombinants, and 1 from subtype C. Crucially, we found that 4 BF recombinant viruses shared a recurrent recombination breakpoint region between positions 4900 and 4924 (relative to the HXB2) that includes 2 gRNA loops, where the RT may stutter. Since these recombinants had independent phylogenetic origin, we argue that these results suggest a possible recombination hotspot not observed so far in BF CRF in particular, or in any other HIV-1 CRF in general. Additionally, 40% of the drug-naïve and 45% of the drug-treated patients had at least 1 raltegravir (RAL) or elvitegravir (EVG) resistance-associated amino acid change, but no major resistance mutations were found, in line with other studies. Importantly, V151I was the most common minor resistance mutation among B, F, and BF *IN* genes. Most codon sites of the *IN* genes had higher rates of synonymous substitutions (*dS*) indicative of a strong negative selection. Nevertheless, several codon sites mainly in the subtype B were found under positive selection. Consequently, we observed a higher genetic diversity in the B portions of the mosaics, possibly due to the more recent introduction of subtype F on top of an ongoing subtype B epidemics and a fast spread of subtype F alleles among the B population.

**Citation:** Iamarino A, de Melo FL, Braconi CT, Zanotto PMdA (2012) BF Integrase Genes of HIV-1 Circulating in São Paulo, Brazil, with a Recurrent Recombination Region. PLoS ONE 7(4): e34324. doi:10.1371/journal.pone.0034324

**Editor:** Darren P. Martin, Institute of Infectious Disease and Molecular Medicine, South Africa

**Received:** October 18, 2011; **Accepted:** February 25, 2012; **Published:** April 2, 2012

**Copyright:** © 2012 Iamarino et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; projects #10/19341-4, 00/04205-6 [VGDN program], 08/58559-5 [AI], 07/01554-9 [FLM], and 09/16740-8 [CTB]). PMAZ holds a CNPq-PQ scholarship. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* E-mail: pzanotto@usp.br

## Introduction

In Brazil, BF recombinants have increasing importance in the HIV epidemic [1,2,3], similar to South America [4], despite subtype C [5] prevalence in the country's south region. To date, 9 BF circulating recombinant forms (CRFs) are worldwide recognized, with 6 of them sampled in Brazil (CRF28\_BF, CRF29\_BF, CRF38\_BF, CRF39\_BF, CRF40\_BF, and CRF46\_BF; www.hiv.lanl.gov). Also, unique recombinant forms (URFs) are commonly found [6,7,8]. Currently, there are 4 different antiretroviral therapies (ARV) for HIV treatment based on: (i) fusion inhibitors that inhibit conformational changes in the gp41 that are essential for fusion; (ii) reverse transcriptase (RT) inhibitors, which may be separated into 2 groups, nucleoside analogs (nucleoside reverse transcriptase inhibitors [NRTIs]) or non-nucleoside analogs (non-nucleoside reverse transcriptase inhibitors [NNRTIs]); (iii) protease (PR) inhibitors; and more recently, (iv) integrase inhibitors such as raltegravir (RAL) [9] and elvitegravir (EVG) [10]. The integrase (*IN*) gene codes for an enzyme responsible for the integration of the full proviral cDNA in the host chromosome. The HIV-1 *IN* is a 32-kDa protein with 288 amino acids (aa) expressed together with a viral PR and RT as the Pol polyprotein and released by PR cleavage during maturation. This enzyme has 3 distinct functional domains: the N-terminal domain (NTD; aa, 1–49), catalytic core domain (CCD; aa, 50–212), and C-terminal domain (CTD; aa,

213–288) [11]. The NTD has a highly conserved zinc-binding HHCC motif that stabilizes its correct folding. In the CCD, it can be found a highly conserved acidic motif, D64, D116, and E152, in all integrases and retrotransposases [11,12]. The last domain, CTD, binds DNA nonspecifically and plays an important role in the integration process [13]. Although, the *IN* gene has become an important target in the ARV therapy, some mutations in the *IN* gene also confer resistance to the integrase inhibitors [14,15].

Brazil started offering free access to HAART treatment in 1996 through the public health system. However, only in the beginning of 2008, RAL was approved for clinical use in patients who had HAART failure. Even though there is a large number of patients under ARV therapy, since the beginning of the epidemic, more than 217,000 people died as a consequence of AIDS, and over 630,000 might be infected by HIV [16]. The *IN* gene in Brazilian patients was previously studied by Passaes et al., 2009 [17] and Arruda et al., 2010 [18], and major resistance mutations were not found in integrase inhibitor-naïve patients, although minor mutations associated with RAL or EVG resistance were common. Nevertheless, neither of these previous works addressed the key aspect of integrase recombination. Our study aimed to sample the genetic diversity of the *IN* gene from the most prevalent subtypes in Brazil, previous to the introduction of RAL in order to evaluate its recombination profiles and drug resistance mutations in a population-genetics context, which is relevant to the epidemic in

the southern cone in particular, and elsewhere in general, once the Brazilian-borne BF CRFs spread abroad [19].

## Methods

### Ethics Statements and Clinical samples

This study was submitted to and was approved by the Ethics Committee on Human Research of the Centro de Referência e Treinamento-DST/AIDS, Sao Paulo, Brazil, and all the patients signed informed consent terms. Patients were selected from those sampled and subtyped for parts of the *gag*, *pol* and *env* regions during the Viral Genetic Diversity Network (VGDN) program [20]. One hundred fifty-seven HIV-1 samples were obtained from drug-naïve and drug-treated individuals; none of which was treated with integrase inhibitors, from different cities of the São Paulo State.

### DNA Extraction, PCR, and Sequencing

DNA was extracted from the infected PBMCs by using QIAamp DNA Blood kit (Qiagen, Germany) according to the manufacturer's instructions and stored at  $-80^{\circ}\text{C}$  until use. The nested-PCR and sequencing were carried out according to Van Laethem et al. (2008) [21] (Table S1). PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Germany). Sequencing reactions were performed using BigDye™ Terminator version 3.0 cycle sequencing (ABI PRISM®; PE Applied Biosystems, Foster City, CA), and the products were analyzed on ABI 3100 automated DNA sequencers (PE Applied Biosystems, United States). Sequence data were edited and assembled with the CodonCode Aligner software (Gene Codes Corporation, United States). The sequences were submitted to GenBank under the accession numbers JN234023 to JN234179.

### Sequence Analysis

To expand our original dataset we included all the *IN* sequences from Brazil available in GenBank. The sequences were analyzed with the jumping profile hidden Markov model (jpHMM) program [22,23], which uses detailed information on polymorphism of the parental populations rather than using individual parental strains and provides detailed information on the reliability of the predicted recombination breakpoints [24]. The phylogenetic analyses were performed using maximum likelihood (ML) as the optimality criterion. The ML searches were performed using GARLI v0.951 (Os×GUI) [25] under the general-time reversible (GTR) model with rates estimated from the data. Only sequences without recombination evidence within the integrase region were used for phylogenetic reconstruction. EVG and RAL resistance-associated mutations and polymorphisms were detected using the Stanford University HIV Drug Resistance Database [26].

### Positive Selection Analysis

Positive selection was determined using the single-likelihood ancestor counting (SLAC) and fixed-effects likelihood (FEL) methods implemented in the HyPhy v2.0 [27]. For the subtype B analysis, a dataset of 324 integrases (126 new sequences and 198 from GenBank); for the subtype F analysis, 80 integrases (17 new sequences and 63 references); and for subtype C analysis, 53 integrase sequences (1 new sequence and 52 references) were compiled. Only results with a *p*-value smaller than 0.05 were considered. We also compared the average rates of synonymous substitutions per site between *IN* genes estimated with SLAC from different subtypes, as an approximation of the neutral rate of change. This is justified because it is estimated at codon site positions that may change with less deleterious impact and make

special sense for the *IN* genes, which are highly conserved and expected to have reduced non-synonymous rates due to purifying selection.

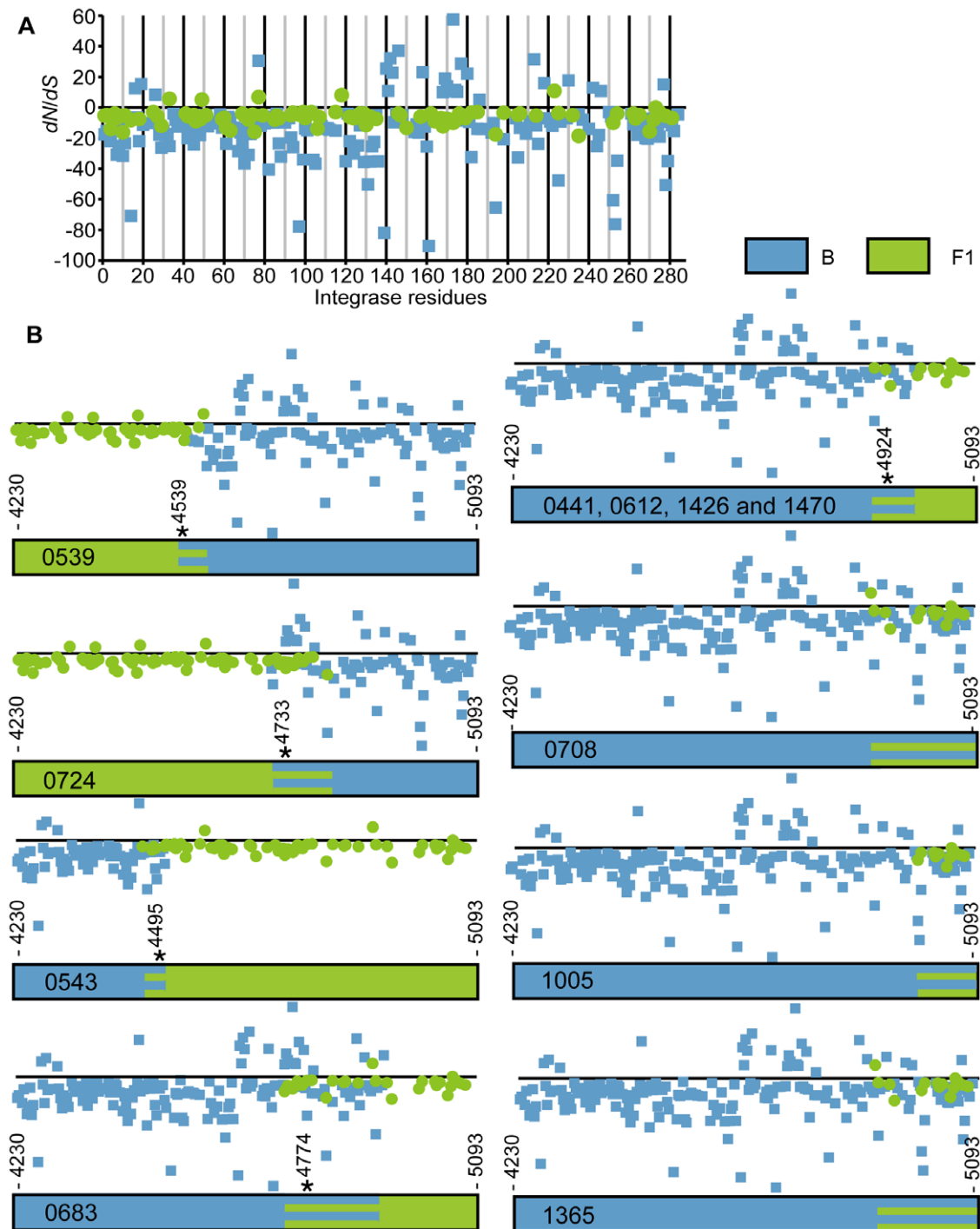
## Results

We examined the almost complete *IN* gene (*p31*; [864 out of 867 bp] genomic positions 4230–5093 in the HXB2 reference) in 157 patients from different cities in the State of São Paulo, Brazil. The jpHMM subtyping analysis revealed that 128 integrases belonged to the subtype B, 17 to the subtype F, 11 to the BF recombinants (Figure 1), and 1 to the subtype C. To further confirm the viral subtype, we compared the integrase subtyping results with those available for other genomic regions. In addition to the recombinant viruses described above (11/157), another 31 patients were identified as carrying a BF recombinant virus, despite not having recombinant *IN* genes. Eight subtype F and 23 subtype B integrases were among the subtype BF genomes.

### The Origin of Integrase Recombinants

Since our results indicated that the vast majority of BF recombinant viruses were not monophyletic (Figure 2), we needed to consider if they had distinct, independent origins as if emerging from distinct co-infection events, despite similarities in their inferred breakpoints. Actually, the high number of URFs, seen in Figure 2, agreed with the reports on a large number of unique BF forms in the Brazilian HIV-1 epidemic [3,6,7]. However, 4 samples shared a recombination breakpoint region between positions 4900 and 4924 (e.g., 0441, 0612, 1426, and 1470) (Figure 1), which could be explained by a shared evolutionary history or maybe by the presence of a recombination hotspot in that particular position in the genome, among other possibilities. To investigate if these recombinants have a common origin, we did a phylogenetic analysis on the first 500 nt (recombination-free segments), including all the subtype B integrases and 8 recombinant integrases (0441, 0612, 0683, 0708, 1005, 1365, 1426, and 1470) that shared this 500-bp region with the subtype B.

Nevertheless, only four of the BF integrase sequences had well defined breakpoints, but a short F portion of around 150 bp-long, making them unsuitable for robust phylogenetic inference on the origin of the F fragments alone. Figure 3 shows that recombinant sequences dispersed all over the tree as if they had independent origins. We found the posterior probability of a possible shared ancestral node for all the recombinants (*i.e.*, testing for monophyly) to be zero on a set of 5000 plausible trees sampled at stationarity during 2 independent Markov chain Monte Carlo (MCMC) runs, each consisting of 20 million steps with the BEAST v1.6.1. Moreover, we also reconstructed the “recombinant” and “non-recombinant” states on the same set of plausible trees by using the symmetric CTMC model for discrete character reconstruction with BEAST and found that all recombinant states had posterior probabilities of 0.99 for independent events of transition from “non-recombinant” to “recombinant” nested within the clades of “non-recombinant” integrases with a posterior probability of 1 (Figure 3). We did not find any other HIV CRF genome with a recombination breaking point at that position. The recurrence of independently acquired breaking points suggests some sort of underlying facilitating process at the molecular level. Crucially, the region where recombination breakpoints had the highest posterior probabilities converged in 2 loops, comprising a highly structured viral genomic RNA (gRNA) region [28] (Figure S1) that is prone to RT stalls and strand transfer during HIV cDNA synthesis involved in HIV recombination [28].

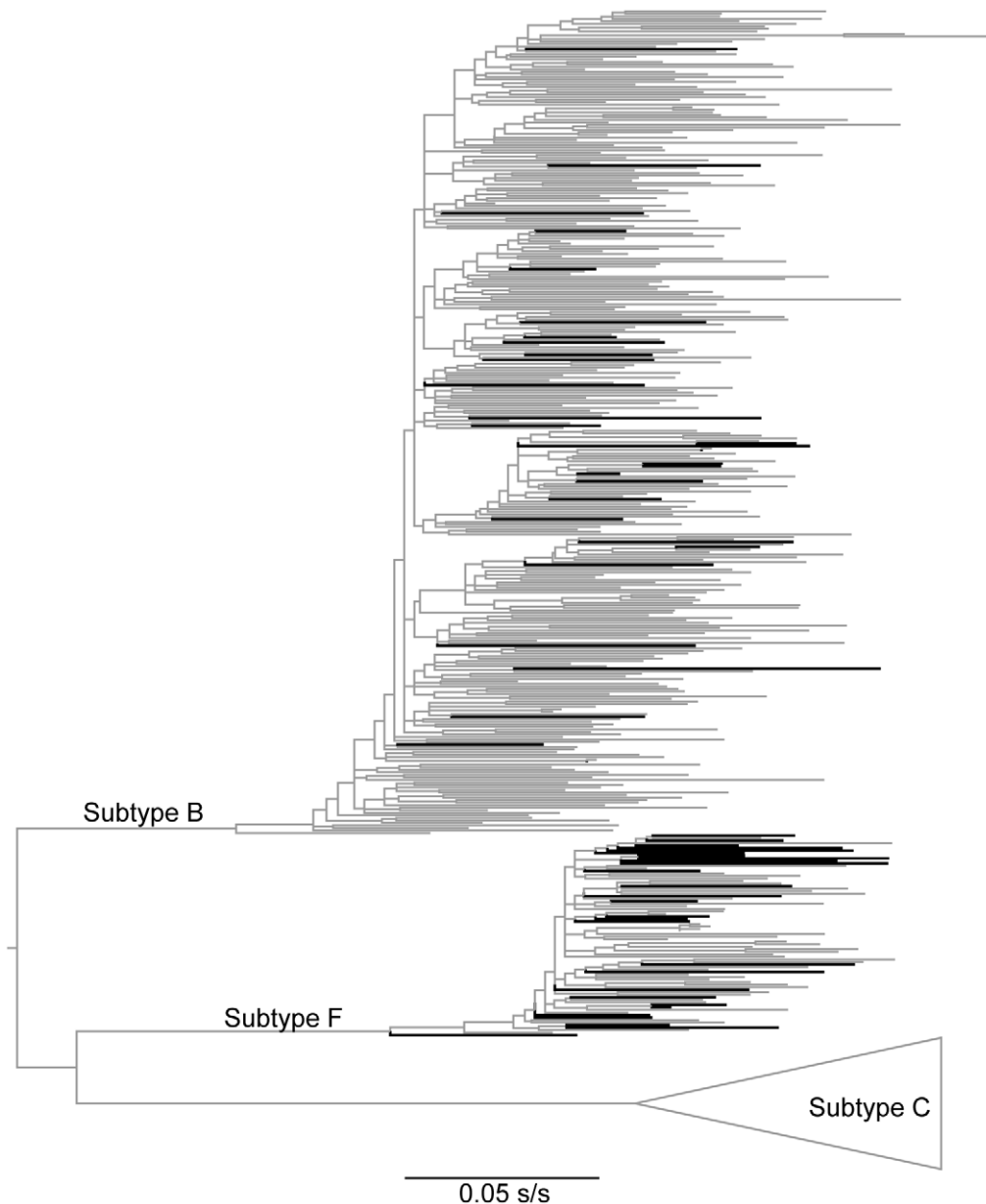


**Figure 1. Recombination profiles and selection.** Plot of the differences between non-synonymous ( $dN$ ) and synonymous ( $dS$ ) rates shown above in the mosaic maps of integrase (864 bp) with nucleotide sites numbered in relation to the HXB2. Rate values per codon site for the subtype B (blue squares) and subtype F (green circles). Blue and green striped portions indicate uncertainty on subtype provenance and usually include the breakpoints (indicated by an asterisk) estimated with jphMM algorithm available online at the GOBICS server (<http://jphmm.gobics.de/jphmm.html>). Samples 0441, 0612, 1365, and 1470 shared the same breakpoint at nucleotide 4924 that falls inside the uncertainty region from 4899 to 4977 bp but had different recombination profiles on parts of their genomes. Sample 0708 also shared a recombination spot starting at 4899 bp, but no breakpoint could be determined. doi:10.1371/journal.pone.0034324.g001

### Relevant Integrase Polymorphisms

One hundred and thirty-two of the 157 patients (84.08%) had already been treated with at least 1 drug regimen (mean, 3; up to 13 distinct drug combinations), 85.6% of which were subjected to

HAART. However, none of them received integrase inhibitors. Forty percent of the drug-naïve and 45.5% of the drug-treated patients had at least 1 RAL or EVG resistance-associated amino acid change, which were values lower than those recently found in



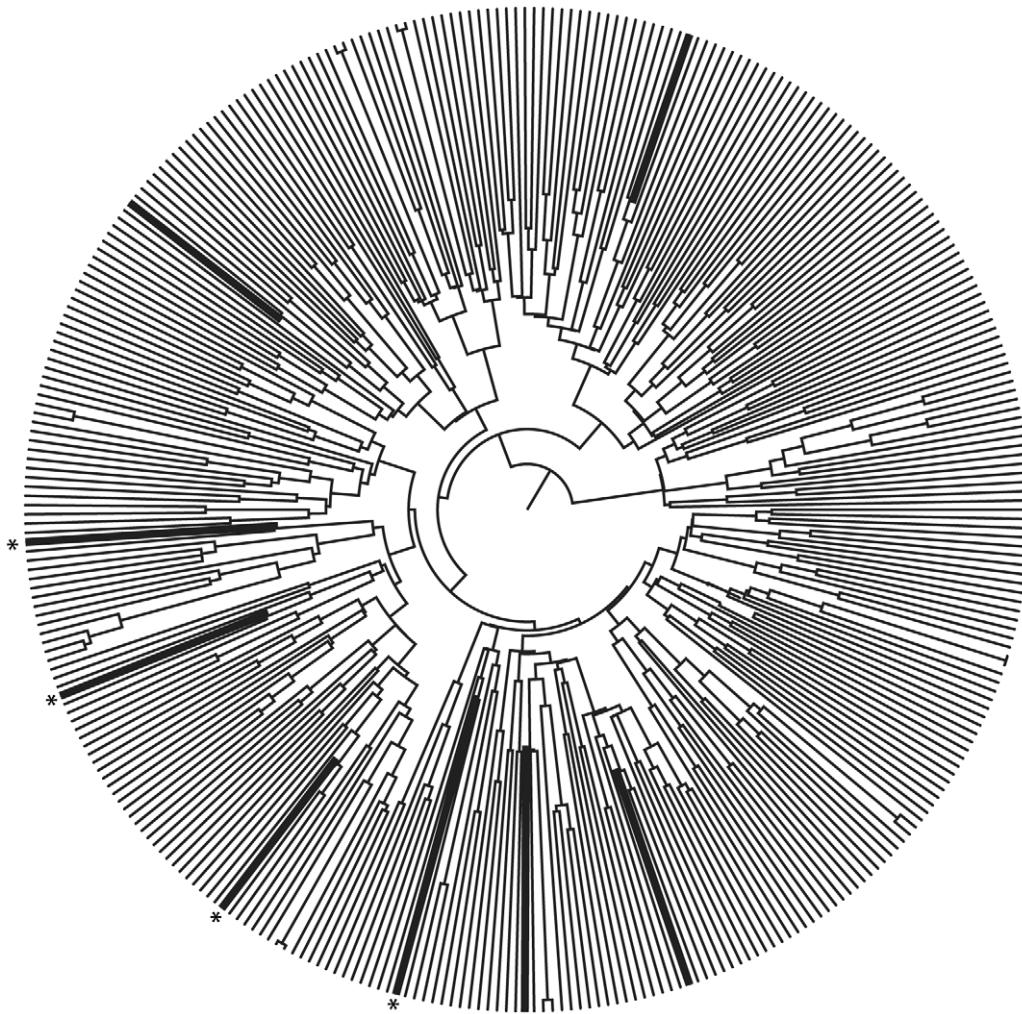
**Figure 2. Integrase phylogeny.** A maximum likelihood tree for 459 non-recombinant integrase sequences from Brazil (145 generated by us and 314 from other studies) inferred with GARLI v0.95. Thick branches indicate lineages known to be recombinants on other regions of their genomes. The tree was midpoint rooted, which agrees precisely with the known relationship among HIV-1 group M subtypes [34]. Subtype C sequences are collapsed for clarity. This tree shows a large number of potential independent origination events of BF recombinants. doi:10.1371/journal.pone.0034324.g002

the State of São Paulo [18]. No major resistance mutations (66I, 92Q, 140A/S, 143R/C, 147G, 148H/R/K, and 155H) were found. Half of the 16 subtype F integrases had minor resistance-associated mutations. V151I was the most frequent and found to be present in 65.2% (5/8) of the cases. The polymorphisms E138D, M154I, and S230N were also found in 1 sample each. Our single subtype C patient was drug-naïve and had no resistance-associated mutation. Likewise, V151I (25%) was the most common minor resistance mutation among the BF recombinant integrases, which also had 1 sample sporting M154I. In the subtype B samples, V151I was also the most frequent change in 23.3% of the patients. Interestingly, unlike the F and BF samples, S230N was the second most common mutation,

present in 8.5% of the subtype B sequences. Other B resistance-associated mutations were L74M (0.8%), T97A (1.6%), E138K (0.8%), M154I (3.9%), M154L (0.8%), E157K (2.3%), G163R (3.1%), and I203M (3.1%). The complete list of polymorphisms is available in Table S2.

### Selection Regimen

Figure 1 shows the difference between non-synonymous ( $dN$ ) and synonymous ( $dS$ ) rates at each codon of the *IN* genes. Positive values indicate an excess of  $dN$  that is indicative of a directional or positive selection, while negative values indicate codons under higher  $dS$  rates, suggestive of purifying or negative selection. The extent of the differences indicates the strength of selection. Based



**Figure 3. Phylogenetic tree using recombination-free segment.** The cladogram summarizes the maximum clade credibility (MCC) tree including all subtype B integrases and the recombinant integrase sequences sharing the 500 bp subtype B segment. It shows reconstructions in “recombinant” and “non-recombinant” state transitions on a set of 5000 plausible trees by using the symmetric CTMC model for discrete character reconstruction with BEAST. Recombinant sequences are shown as thick branches, while asterisks at the tips of the MCC tree show sequences that share a breakpoint at position 4924. Branch lengths only help the visualization of transition events and are not proportional to either quantity of events or substitution per sites.  
doi:10.1371/journal.pone.0034324.g003

only on  $dS$  values, it can be observed higher genetic diversity in the B portions of the mosaics, which is in agreement with the notion of a more recent introduction of the F on top of a B epidemic, possibly in concert with a fast spread of F alleles among the B population. The fact that most of the codon sites of the integrase had higher  $dS$  agrees with its functional relevance for viral replication revealed by the strong negative selection on deleterious changes. Nevertheless, several codon sites mainly in the subtype B were found under positive selection using the SLAC method, despite the fact that the integrase is relatively conserved compared to other HIV-1 genes. Some of the positively selected sites, such as residue 72 in the subtype B and F integrases and residues 151 and 201 in the subtype B, were further confirmed using FEL. Moreover, the subtype F integrases had 3 additional positively selected sites at the CCD. Subtype B, on the other hand, presented many positively selected sites, including residues with important biological functions, such as the active site loop (aa, 140) [11]. Interaction with human lens epithelium-derived growth factor (LEDGF/p75; aa: 161, 166, 168, and 171) [29,30], DNA binding

(aa, 163) [30], and the nuclear import signal (aa, 165 and 166) [31]. At the CTD, positive selection was found only in the subtype B integrase, including at the DNA binding sites (aa: 230, 234, and 284) [29]. At the NTD, the subtype B and subtype F had 2 positively selected sites (at residues: 17 and 24 and at residues 31 and 45, respectively). Furthermore, the most common BF integrase recombination profile (breakpoint at residue 232) preserved the positively selected NTD and CCD domains, and the subtype F conserved the DNA binding site at CTD. We also found that the B integrases ( $dS_B = 14.42 \pm 1.88$ ) had a 4.74 times higher rate of change estimated at silent sites ( $dS$ ) compared to F integrases ( $dS_F = 3.04 \pm 0.44$ ) ( $dS_B > dS_F$ , with a probability of  $5.9E-28$  of being the same value given a Student's paired  $t$ -test, with a 2-tailed distribution). This huge difference entails a larger population size and possibly an earlier entry of the subtype B in our epidemics. The complete list of positively selected sites is available in Table S3.

## Discussion

We found that the *IN* gene of the HIV-1 circulating in the epidemics in the São Paulo State displayed a diverse plethora of BF recombinants. Our findings are relevant since not much is known about the diversity of circulating viruses and their background resistance-associated polymorphisms in general. The absence of major integrase resistance mutations in our integrase inhibitor-naïve patients, 79% of which were under ARV drugs such as RT and PR inhibitors, is in line with other studies [17,18] and justifies the use of this therapeutic strategy. None of the 11 recombinant integrases we found shared the same breakpoints with the already described BF CRFs from Brazil, although CRF47\_BF from Spain has a breakpoint close to those found in 0441, 0612, 1426, and 1470. These results show the importance of sequencing different genomic regions for improving HIV subtyping and give information on the potential of the underestimation of the real diversity of BF recombinants in our epidemics. Moreover, we tested the hypothesis that the recombinant genes originated from a single common ancestral recombination event (testing for monophyly) by calculating the posterior probability (that was zero) of monophyly on a set of 5000 plausible trees sampled at stationarity during 2 independent MCMC runs. We also checked these results by reconstructing the “recombinant” and “non-recombinant” states on the same set of plausible trees. This exercise also reinforced (with posterior probability of 1) the hypothesis of an independent origin of “recombinant” from “non-recombinant” states within the clades of “non-recombinant” integrases. Critically, both independent approaches provided strong statistical support for the notion that the recombinant sequences we found originated from distinct recombination events. That being the case and given the fact that the recurrent recombination region (Figure S2) fell within a secondary gRNA structural feature known to propitiate recombination, we argue that it may indicate a potential recombination hotspot, which was not previously reported for HIV-1 in general and BF recombinants in particular. Nevertheless, *IN* is a very conserved region and therefore prone to recombination [32,33]. Moreover, we cannot be sure of the extent to which our estimates could have been influenced by the recombination itself. This is because recombinants with common origin could differentiate later by distinct subsequent recombination events with divergent parental strains

## References

- Bello G, Guimaraes ML, Morgado MG (2006) Evolutionary history of HIV-1 subtype B and F infections in Brazil. *AIDS* 20: 763–768.
- de Souza AC, de Oliveira CM, Rodrigues CL, Silva SA, Levi JE (2008) Short communication: Molecular characterization of HIV type 1 BF pol recombinants from Sao Paulo, Brazil. *AIDS Res Hum Retroviruses* 24: 1521–1525.
- Sanabani S, Kleine Neto W, Kalmar EM, Diaz RS, Janini LM, et al. (2006) Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. *Infect Genet Evol* 6: 368–377.
- Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, et al. (2010) The role of recombination in the emergence of a complex and dynamic HIV epidemic. *Retrovirology* 7: 25.
- Gonsalez CR, Alcalde R, Nishiya A, Barreto CC, Silva FE, et al. (2007) Drug resistance among chronic HIV-1-infected patients naive for use of anti-retroviral therapy in Sao Paulo city. *Virus Res* 129: 87–90.
- De Sa Filho DJ, Sucupira MC, Caseiro MM, Sabino EC, Diaz RS, et al. (2006) Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res Hum Retroviruses* 22: 1–13.
- Sa Filho DJ, Sanabani S, Diaz RS, Munerato P, Brunstein A, et al. (2005) Analysis of full-length human immunodeficiency virus type 1 genome reveals a variable spectrum of subtypes B and f recombinants in Sao Paulo, Brazil. *AIDS Res Hum Retroviruses* 21: 145–151.
- Sanabani S, Neto WK, de Sa Filho DJ, Diaz RS, Munerato P, et al. (2006) Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res Hum Retroviruses* 22: 171–176.
- Grinsztejn B, Nguyen BY, Katlama C, Gatell JM, Lazzarin A, et al. (2007) Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial. *Lancet* 369: 1261–1269.
- DeJesus E, Berger D, Markowitz M, Cohen C, Hawkins T, et al. (2006) Antiviral activity, pharmacokinetics, and dose response of the HIV-1 integrase inhibitor GS-9137 (JTK-303) in treatment-naïve and treatment-experienced patients. *J Acquir Immune Defic Syndr* 43: 1–5.
- Engelman A, Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J Virol* 66: 6361–6369.
- Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* 12: 2331–2338.
- Pommier Y, Johnson AA, Marchand C (2005) Integrase inhibitors to treat HIV/AIDS. *Nat Rev Drug Discov* 4: 236–248.
- Lataillade M, Chiarella J, Kozal MJ (2007) Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. *Antivir Ther* 12: 563–570.
- Malet I, Delelis O, Valantin MA, Montes B, Soulie C, et al. (2008) Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob Agents Chemother* 52: 1351–1358.
- Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde, Programa Nacional de DST e AIDS (2010) *Boletim Epidemiológico*. Ano VII n0 01 01/12/2010 ed: Brasília. 32 p.

(i.e., second-generation recombinants). We also observed a 4.7-times higher accumulation of diversity in the B subtypes compared to F subtypes in the *IN* genes. Since there was a reduction in the number of positively selected sites in the F regions and given that the most positively selected sites were detected at the B portion of recombinant forms, we argue that this higher genetic diversity could be explained by a more recent introduction of F on top of a B epidemics and a fast spread of F alleles among the larger B population. It remains to be investigated if the advantage conferred by the F portions may have to do with the intrinsic F integrase functionality or adaptive changes acquired following coinfection.

## Supporting Information

**Figure S1 Breakpoint region in secondary RNA structure.** Secondary structure of the HIV genomic RNA (gRNA) numbered according to HXB2. The red region marks the recombination hotspot from bases 4900 to 4924. Adapted from Watts J. M. et al. (2009). (TIF)

**Figure S2 Recombination posterior probabilities.** Posterior probabilities according to jpHMM of the subtypes for patients 0441 (A), 0612 (B), 1426 (C), and 1470 (D). (TIF)

**Table S1** Primers used for nested amplification of the HIV-1 IN region from according to Van Laethem K.V. et al. (2008). (DOC)

**Table S2** Complete list of polymorphisms found at integrase sequences. Minor resistance mutations are shown in red. (DOC)

**Table S3** Positively selected sites in the HIV-1 integrase gene coding region. (DOC)

## Author Contributions

Conceived and designed the experiments: AI FLM CTB PMAZ. Performed the experiments: AI FLM CTB. Analyzed the data: AI FLM CTB PMAZ. Contributed reagents/materials/analysis tools: AI FLM CTB PMAZ. Wrote the paper: AI FLM CTB PMAZ.

17. Passaes CP, Guimaraes ML, Bello G, Morgado MG (2009) Near full-length genome characterization of HIV type 1 unique BC recombinant forms from Southern Brazil. *AIDS Res Hum Retroviruses* 25: 1339–1344.
18. Arruda LB, Fonseca LA, Duarte AJ, Casseb J (2010) Genetic diversity on the integrase region of the pol gene among HIV type 1-infected patients naive for integrase inhibitors in Sao Paulo City, Brazil. *AIDS Res Hum Retroviruses* 26: 105–107.
19. Bruselles A, Rozera G, Bartolini B, Prosperi M, Del Nonno F, et al. (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.
20. Pardini MI, Jamal LF, Durigon EL, Massad E, Perez JF, et al. (2008) Boosting virology in Brazil. *PLoS Biol* 6: e57.
21. Van Laethem K, Schrooten Y, Covens K, Dekeersmaecker N, De Munter P, et al. (2008) A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J Virol Methods* 153: 176–181.
22. Schultz AK, Zhang M, Leitner T, Kuiken C, Korber B, et al. (2006) A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes. *BMC Bioinformatics* 7: 265.
23. Zhang M, Schultz AK, Calef C, Kuiken C, Leitner T, et al. (2006) jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res* 34: W463–W465.
24. Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, et al. (2009) jpHMM: Improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res* 37: W647–W651.
25. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [Doctoral Thesis]. Austin: The University of Texas.
26. Shafer RW (2006) Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 194 Suppl 1: S51–58.
27. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Bio Evol* 23: 1891–1901.
28. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460: 711–716.
29. Ceccherini-Silberstein F, Malet I, D'Arrigo R, Antinori A, Marcelin AG, et al. (2009) Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev* 11: 17–29.
30. Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468: 326–329.
31. Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, et al. (2001) HIV-1 infection requires a functional integrase NLS. *Mol Cell* 7: 1025–1035.
32. Magiorkinis G, Paraskevis D, Vandamme AM, Magiorkinis E, Sypsa V, et al. (2003) In vivo characteristics of human immunodeficiency virus type 1 intersubtype recombination: determination of hot spots and correlation with sequence similarity. *J Gen Virol* 84: 2715–2722.
33. Thomson MM, Najera R (2005) Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev* 7: 210–224.
34. McCutchan FE (1999) Global Diversity in HIV. In: Crandall KA, ed. *The evolution of HIV*. Baltimore: The Johns Hopkins University Press. pp 41–104.