Genetic Variability of Human Immunodeficiency Virus-1 in Bahia State, Northeast, Brazil: High Diversity of HIV Genotypes

Joana Paixão Monteiro,1 Luiz Carlos Junior Alcantara,1,2 Tulio de Oliveira,3 Antonio Marcos Oliveira,1 Marco Antônio Gomes Melo,1,2 Carlos Brites,4 and Bernardo Galvão-Castro1,2*

1Laboratório Avançado de Saúde Pública, CPqGM/FIOCRUZ, Salvador, Bahia, Brazil
2Escola Bahiana de Medicina e Saúde Pública, EBMS, Salvador, Bahia, Brazil
3South African National Bioinformatics Institute, Cape Town, South Africa
4Universidade Federal da Bahia, Salvador, Bahia, Brazil

The HIV-1 genetic variability in Bahia state, Brazil, was investigated. DNA samples from 229 and 213 HIV-1-infected individuals were analyzed using the heteroduplex mobility assay (HMA) in gag and env fragments, respectively. One hundred seventy-five samples were characterized in both genes. Thirty-two subtype F and BF recombinant viruses were sequenced and analyzed by phylogenetic methods. The combination of HMA and sequencing results showed that seven different HIV-1 genotypes comprised this sample: 147 (84%) B/B, 4 (2.3%) F/F, 3 (1.7%) B/F, 1 (0.6%) F/B, 1 (0.6%) F/D, 1 (0.6%) BF/F, and 18 (10.3%) BF/B. A significant divergence was observed between these two techniques results (84.4%). This is explained by the low accuracy of the HMA for detecting recombinant viruses. These recombinants were unrelated to CRF12, while two sequences were related to CRF28 and CRF29. Nineteen BF mosaics shared the same gag breakpoint. In conclusion, the use of HMA may be inappropriate in regions where different subtypes are co-circulating. Subtype B is the most common genotype, however, an increased prevalence (13.1%) of different BF variants and a potentially new CRF suggest that recombination is occurring frequently in Bahia. These viruses were associated with women infected heterosexually. Finally, this study identified the presence of an F/D recombinant HIV-1 in Brazil. J. Med. Virol. 81:391–399, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: HIV-1; epidemiology; phylogeny; Brazil; Bahia

INTRODUCTION

The human immunodeficiency virus (HIV) is characterized by high genetic variability and is classified into different types, groups, subtypes, and circulating recombinant forms (CRFs) [Mccutchan et al., 2000; Casado et al., 2005; Thomson et al., 2005; De Sa Filho et al., 2006], each presenting distinct distribution in the world. Moreover, HIV-1 genotypes show different dispersion among the different risk groups and transmission routes [Weniger et al., 1994; Kunanusont et al., 1995; Burke and McCutchan, 1996]. Nevertheless, the association between this genetic variability and viral features as transmission, infectivity, and pathogenesis is not elucidated to date.

Subtype B is the predominant HIV-1 genotype in Brazil. However, an increasing number of non-B and recombinants infections are being identified [Morgado et al., 2002; Eyer-Silva et al., 2007; Monteiro et al., 2007]. Within the Brazilian vast territory, different subtype distribution patterns have been reported among its five geographic regions. The southern and southeastern regions have been largely characterized in regard to HIV-1 molecular epidemiology. However, little information is available about other regions of the country. Particularly, in Bahia state, few previous studies have focused on the prevalence of HIV-1 subtypes [Couto-Fernandez et al., 1999; Pedroso et al., 2007] which considered only short fragments of the HIV genome, thus, probably under estimating the presence of mosaic virus.

Grant sponsor: Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB); Grant sponsor: Brazilian Ministry of Health STD/AIDS Program; Grant sponsor: CNPq.

*Correspondence to: Bernardo Galvão-Castro, Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Munique Fundação Oswaldo Cruz, Rua Waldemar Falcão 121, Candeal, Salvador, Bahia 40296-610, Brazil.
E-mail: bgalvao bahia.fiocruz.br
Accepted 4 November 2008
DOI 10.1002/jmv.21414
Published online in Wiley InterScience (www.interscience.wiley.com)
Bahia is the fourth most populous Brazilian state and presents the highest number of notified cases of HIV/AIDS in the Northeast region. In this part of the country, economic and education privation, sexual tourism, and prostitution can be related to an increase in the incidence of AIDS and the introduction of new genetic forms is expected. Therefore, the identification of these variants is of great importance for HIV surveillance as well as for vaccine development and efficacy. This study was conducted in order to access the HIV genetic variability in Salvador, the capital of Bahia, to determine the prevalence of subtypes and to verify the association with groups and transmission routes. Also, the current molecular epidemiological profile of the HIV population in this area was compared with previous data from 10 years ago. Thus, besides its potential implications for therapeutic strategies at a local level, this study contributes to the understanding of the evolution of HIV.

This study has received official and ethical approval of the Centro de Pesquisa Gonçalo Moniz, Fiocruz, Bahia.

**METHODS**

**Study Population**

As part of a sectional study in 2002, blood samples from 261 HIV-1-seropositive subjects, unrelated epidemiologically and followed regularly at Professor Edgard Santos University Hospital, Salvador, Bahia, were obtained after informed consent. The samples were sent to the Advanced Laboratory of Public Health (LASP)/CPqGM/FIOCRUZ for further processing. Peripheral blood mononuclear cells (PBMCs) were separated by centrifugation on Ficoll-hypaque gradient. DNA was extracted from PBMCs using the Qiagen extraction kit and submitted to phylogenetic analysis using the NJ tree as the starting material, including its optimized parameters. The likelihood ratio test (RT) method was used to calculate statistical support for the branches: P < 0.001 (highly significant **) and P < 0.005 (significant *).

**Genotyping**

Semi-nested PCR was conducted for individual amplification of gag and env fragments. The heteroduplex mobility assay (HMA) was based on gag and env fragments amplified by PCR. H1G777/H1P202 and H1Gag1584/G17 were used as outer and inner primer sets for gag, while for env, ED5/ED12 and ED31/ED33 were used as outer and inner primers, respectively. The PCR conditions and the HMA subtyping were carried out as described previously [Delwart et al., 1993; Heyndrickx et al., 2000]. Samples classified as subtype F in HMAgag and/or HMAenv had further PCR amplification. The gag p17 (~1,150 bp) nested PCR protocol was carried out as described previously [Guimarães et al., 2002]. The same HMA primer set (~550 bp) or, alternatively, EST7/ES8 (~700 bp) were used for env amplification.

**Sequencing and Analysis**

PCR products were purified in Qiagen columns (Qiagen) and sequenced in an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) using the Big Dye Terminator kit (Applied Biosystems). Sequences were assembled using CLUSTAL X software [Thompson et al., 1997] and edited manually in GENEDOC [Nicholas et al., 1997]. The gag and env fragments corresponded to positions 952–1958 and 6845–7924 relative to HXB2 genome, respectively. All the sequences were checked for contamination by BLAST search against the HIV-1 sequences database and among themselves. Neighbor-joining (NJ) and maximum likelihood (ML) trees were reconstructed in the PAUP* 4.0b10 software [Swofford, 1999], using the appropriate nucleotide substitution model implemented in the Modeltest software [Posada and Krandall, 1998]. The reliability of each cluster was determined using 1,000 bootstrap replicates. For the ML tree, a heuristic search was undertaken with a subtree-pruning-regrafting branch swapping algorithm using the NJ tree as the starting material, including its optimized parameters. The likelihood ratio test (RT) method was used to calculate statistical support for the branches: P < 0.001 (highly significant **) and P < 0.005 (significant *).

**RESULTS**

In this sectional study, HMA analyses were conducted in order to identify the HIV-1 subtype of 155 men (59.4%) and 106 (40.6%) women. Out of these, 228 and 208 samples were individually analyzed in gag and env fragments, respectively. Based on gag analysis, 202 (88.6%) samples were classified as subtype B, 25 (11.0%) as subtype F, and 1 (0.4%) as subtype D, while in env
region, 204 (98.1%) samples were classified as B, 3 (1.4%) as F, and 1 (0.5%) as D. Subtype F was at least 3 times more frequent in \textit{gag} than in \textit{env} gene. One hundred seventy-five samples were characterized in both genomic regions; pure subtype B was assigned for 143 (81.7%) samples and pure subtype F for 1 (0.6%). Three recombinant forms were identified: 6 (3.4%) B\textit{gag}/F\textit{env}, 24 (13.7%) F\textit{gag}/B\textit{env}, and 1 (0.6%) D\textit{gag}/B\textit{env}. Concerning the association between HIV-1 subtypes and gender, 64% (16) of the samples classified as subtype F in \textit{gag} were obtained from women, while 63.4% (128) of the B samples were isolated from men. In \textit{env}, all three F samples represented men and 40.7% (83) of the B samples represented women. Regardless of the genomic segment analyzed, subtype B was the most prevalent among both genders.

In attempt to characterize the subtype F viruses circulating in Brazil, \textit{gag} (952–1958 relative to HXB2) and \textit{env} (6845–7324 relative to HXB2) fragments of the 32 samples classified as subtype F in the \textit{gag} and/or the \textit{env} HMA were sequenced and submitted to phylogenetic analyses (Fig. 1). Some samples are not shown in the trees, either because they had shorter \textit{gag} sequences, were amplified with different \textit{env} primers or have not been amplified in \textit{env} region. In the NJ analysis of \textit{env}, 18 sequences clustered as subtype B, 3 as subtype F, and 1 as subtype D. In \textit{gag} analysis, 7 sequences grouped with subtype B references, 6 sequences grouped within the subtype F reference cluster, and 18 sequences formed a separated but closely related to the F group (bootstrap = 70). Therefore, these sequences were examined further using the Bootscanning recombination tool. All these sequences and one other smaller sequence (109) presented a recombination pattern between subtypes B and F in the \textit{gag} fragment (Fig. 2a), with the crossover site located between positions 1370 and 1410 (relative to HXB2). In order to determine if these BF recombinants were phylogenetic related, the exact recombination point was sought through the inspection of subtype signature nucleotides in comparison to the Los Alamos reference set with the GENEDOC software. All the BF recombinant sequences shared the same breakpoint, located at p24 (position 1398 in HXB2). NJ and ML analyses of the individual fragments using the appropriate evolutionary model confirmed the Bootscanning recombination results (Fig. 2b).
The HMA and phylogenetic results (Table I) were compared. The following discordance was found: based on gag, two samples classified as subtype F (393 and 522) and one as subtype D (168) with HMA were indeed subtype B, while three samples classified as subtype B grouped within the subtype F (29 and 140) and BF (524) clusters. Based on env, two samples classified as subtype B with HMA were actually one subtype F (23) and one subtype D (240) in the phylogenetic tree and one classified as subtype F was inside the subtype B group (69).

Thus, considering the combined HMA and sequencing analyses (Table I), the 175 samples were classified in
seven different HIV-1 genotypes based on gag/env genomic regions: 147 (84%) B/B, 4 (2.3%) F/F, 3 (1.7%) B/F, 1 (0.6%) F/B, 1 (0.6%) BF/F, and 18 (10.3%) BF/B. Table I summarizes the subtypes and clinical-epidemiological data of the 28 individuals infected with subtype F or F recombinant viruses. Overall, 67.9% were women, 79.2% reported heterosexual behavior, 8.3% were intravenous drug users, and 2 women (8.3%) were infected through blood transfusion. One man reported homosexual relation as the transmission route. Mean age was 35 years, median CD4 count was 303 cells/ml (interquartile range = 349) and median viral load was 1,600 copies/ml (interquartile range = 10,000). Most of the patients (83.3%) did not present symptoms typical of AIDS status.

To further characterize the intra-subtype relationships of the viruses circulating in this population and to analyze the relationships between new and previously described BF recombinants, phylogenetic trees were constructed using Los Alamos reference sequences of subtypes B, F, and D and other pure-subtype and recombinant sequences from South America (Fig. 3). In the gag full-segment analysis (Fig. 3a), like in the subtyping tree (Fig. 1), the 18 recombinant sequences clustered as a monophyletic group outside the subtype F1 cluster together with other BF recombinants described previously including sequences of two CRFs identified in southeastern Brazil [De Sa Filho et al., 2005, 2006]. However, two different monophyletic sub-clusters can be observed within this BF cluster: one formed by 16 of the recombinant sequences identified in this study and other BF recombinant sequences from Brazil and Argentina and the other formed by two of the new recombinant samples, CRF28 and CRF29 sequences. Moreover, inside the former cluster, four sequences (454, 524, 152, and 114) formed a separate cluster supported by a bootstrap of 90. The same topology for this group was observed in the trees based on individual subtype fragments of gag (Fig. 2b). None of the recombinants were closely related to the CRF12, although one F sequence (532) was in both inter- and intra-subtype (Fig. 3) analyses of gag. The intra-subtype phylogenetic analysis of the envelope sequences (Fig. 3b) showed that subtype B sequences clustered inside two major groups. The first one was represented by all the CRF28 and CRF29 sequences in addition to five of the new sequences, all of which presenting the BF recombination pattern in gag. The second group was formed by nine samples obtained in this study, all the other sequences from Bahia [Couto-Fernandez et al., 1999] and other B sequences from other parts of Brazil and from other countries. All of

<table>
<thead>
<tr>
<th>Sample</th>
<th>gag</th>
<th>env</th>
<th>Gender</th>
<th>Age (year)</th>
<th>Diagnosis</th>
<th>Transmission route</th>
<th>Viral load (copies/ml)</th>
<th>CD4 (cells/ml)</th>
<th>AIDS symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>38</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Y</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>38</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N</td>
</tr>
<tr>
<td>67</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>Ht</td>
<td>1996</td>
<td>1,600</td>
<td>176</td>
<td>N</td>
</tr>
<tr>
<td>89</td>
<td>BF</td>
<td>B</td>
<td>M</td>
<td>39</td>
<td>1997</td>
<td>Ht</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>90</td>
<td>BF</td>
<td>B</td>
<td>M</td>
<td>39</td>
<td>1997</td>
<td>Ht</td>
<td>4,700</td>
<td>85</td>
<td>N</td>
</tr>
<tr>
<td>109</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>Ht</td>
<td>2000</td>
<td>&lt;40</td>
<td>157</td>
<td>N</td>
</tr>
<tr>
<td>114</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>23</td>
<td>NA</td>
<td>1996</td>
<td>250,000</td>
<td>65</td>
<td>N</td>
</tr>
<tr>
<td>152</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>28</td>
<td>1997</td>
<td>Blood</td>
<td>3,200</td>
<td>694</td>
<td>N</td>
</tr>
<tr>
<td>155</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>36</td>
<td>1996</td>
<td>Ht</td>
<td>8,700</td>
<td>210</td>
<td>Y</td>
</tr>
<tr>
<td>173</td>
<td>BF</td>
<td>B</td>
<td>M</td>
<td>25</td>
<td>1999</td>
<td>Ht</td>
<td>450,000</td>
<td>233</td>
<td>N</td>
</tr>
<tr>
<td>184</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>23</td>
<td>1996</td>
<td>Ht</td>
<td>250,000</td>
<td>65</td>
<td>Y</td>
</tr>
<tr>
<td>185</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>Ht</td>
<td>2001</td>
<td>1,700</td>
<td>393</td>
<td>N</td>
</tr>
<tr>
<td>188</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>Ht</td>
<td>2002</td>
<td>10,000</td>
<td>303</td>
<td>N</td>
</tr>
<tr>
<td>201</td>
<td>BF</td>
<td>B</td>
<td>M</td>
<td>35</td>
<td>NA</td>
<td>1996</td>
<td>66,000</td>
<td>156</td>
<td>N</td>
</tr>
<tr>
<td>217</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>41</td>
<td>2000</td>
<td>IDU</td>
<td>&lt;40</td>
<td>157</td>
<td>N</td>
</tr>
<tr>
<td>239</td>
<td>B</td>
<td>F</td>
<td>F</td>
<td>30</td>
<td>1999</td>
<td>Blood</td>
<td>&lt;80</td>
<td>223</td>
<td>N</td>
</tr>
<tr>
<td>240</td>
<td>B</td>
<td>F</td>
<td>F</td>
<td>38</td>
<td>1998</td>
<td>Ht</td>
<td>110</td>
<td>246</td>
<td>NA</td>
</tr>
<tr>
<td>250</td>
<td>B</td>
<td>F</td>
<td>F</td>
<td>NA</td>
<td>NA</td>
<td>IDU</td>
<td>61,000</td>
<td>407</td>
<td>N</td>
</tr>
<tr>
<td>332</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>NA</td>
<td>NA</td>
<td>2001</td>
<td>1,700</td>
<td>393</td>
<td>N</td>
</tr>
<tr>
<td>414</td>
<td>BF</td>
<td>B</td>
<td>M</td>
<td>39</td>
<td>2002</td>
<td>Ht</td>
<td>10,000</td>
<td>303</td>
<td>N</td>
</tr>
<tr>
<td>415</td>
<td>B</td>
<td>F</td>
<td>F</td>
<td>NA</td>
<td>NA</td>
<td>Ht</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>454</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>NA</td>
<td>Ht</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>500</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>39</td>
<td>NA</td>
<td>Ht</td>
<td>Ldl</td>
<td>612</td>
<td>NA</td>
</tr>
<tr>
<td>524</td>
<td>BF</td>
<td>F</td>
<td>F</td>
<td>35</td>
<td>NA</td>
<td>Ht</td>
<td>Ldl</td>
<td>549</td>
<td>N</td>
</tr>
<tr>
<td>532</td>
<td>F</td>
<td>B</td>
<td>F</td>
<td>35</td>
<td>1996</td>
<td>Ht</td>
<td>Ldl</td>
<td>550</td>
<td>N</td>
</tr>
<tr>
<td>547</td>
<td>BF</td>
<td>B</td>
<td>F</td>
<td>NA</td>
<td>NA</td>
<td>Ht</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

B*, Brazilian subtype B; NA, not available; M, male; F, female; Ht, heterosexual; Hm, homosexual; IDU, intravenous drug user; Ldl, lower than detectable levels; Y, yes; N, no.

gag and env subtypes were defined based on phylogenetic analyses. For the samples that did not have enough material for the sequencing, results presented here are based on HMA.
these nine subtype B sequences were BF recombinants in the \textit{gag} fragment. Concerning the intra-subtype F relationships, the three new envelope sequences were dispersed among the vast reference group; however, none of them was closely related to the two other subtype F sequences from Bahia identified previously neither to the subtype F sequence of CRF12. The \textit{env} subtype D sequence identified here clustered inside the subtype D group, however, this sequence was not related to any of the other sequences of subtype D identified so far in Brazil as well as to any other African sequence used as reference. The genetic distances for the subtypes and recombinant clusters are shown in the trees.

Diverse unique BF recombinant genomes as well as two CRFs have been described in other geographic regions of Brazil [De Sa Filho et al., 2005, 2006; Sanabani et al., 2006]. Thus, the genetic structures of such viruses were compared to the recombination patterns from Bahia (Fig. 4). In the \textit{gag} and \textit{env} concatenated tree, 19 new viruses had both genome fragments available. Out of these, all BF_{gag}/B_{env} and F_{gag}/B_{env} recombinants involved in this analysis formed a unique monophyletic group (bootstrap = 73). Again, two different recombinant clusters, with significant bootstrap values (84 and 90, respectively), were observed: one formed by CRF28 and CRF29 sequences in addition to two of the samples and the other composed by nine new recombinants and one recombinant from Brazil identified previously. The other BF recombinant pattern virus (415 = B_{gag}/F_{env}) clustered closer to subtype B sequences, however, outside the group. The FD recombinant sequence did not cluster with any pure-subtype or recombinant sequence.

\section*{DISCUSSION}

The identification of HIV-1 genotypes represents an important tool for molecular epidemiology studies, helping the evaluation of changes in the HIV epidemic and also contributing to the understanding of the determinants of virus evolution. The HIV genetic diversity and distribution are well documented for the southern and southeastern regions of Brazil; however, few reports have focused on the remaining geographic regions. The aim of this study was to evaluate the HIV-1 subtype distribution in Salvador, the capital of Bahia state and the third most populous city in the country. With almost 3 million people living in the city, 13,750 are estimated to be infected with HIV in Salvador [Dourado et al., 2007]. The HMA testing was used to screen the subtypes in 261 HIV-positive samples since this technique has been described as a useful tool to determine HIV subtypes [Fereira et al., 2004; Rios et al., 2005; Cabral et al., 2006]. However, when the sequences of 32 subtypes F and BF recombinants were analyzed, a significant proportion of divergence was found between these methods (84.4\%). These contrasting results can be attributed to the limited ability of the HMA to
distinguish BF recombinants as 19 intra-gag BF recombinants were identified as either subtype F or B isolates. Therefore, the use of the HMA may be inappropriate to identify HIV genotypes in regions where different subtypes co-circulate and where the presence of recombinant viruses is expected. In this context, Buonaguro et al. [2005] had proposed a modified version of HMA in which instead of being used for subtype classification, the technique is used to screen HIV-1 isolates divergent from the local predominant subtype by testing samples only against the reference standard representing that predominant subtype. Eighty-four percent of the studied samples were pure B, 2.3% were pure F, 13.1% were represented by different patterns of subtypes B and F recombinants, and 0.6% (one sample) was an F/D recombinant. Subtype C was not represented in this population. The differences in subtypes B and BF recombinants prevalences compared to the previous data (92% and 5.5%, respectively) [Couto-Fernandez et al., 1999] can be attributed to two different factors: (1) an increase in the circulation of BF recombinants and (2) a previous sub-estimation of the real prevalence of BF recombinants since only one genetic region was analyzed in that report. On the other hand, both present and previous studies reported a similar lower rate of pure subtype F (1.6%) compared to BF recombinants in Bahia. The 2.3% of subtype F found here are also lower than the prevalence found for the southeastern region, which varied from 3.7% to 9% [Guimarães et al., 2002; Couto-Fernandez et al., 2005; De Sa Filho et al., 2005; Barreto et al., 2006; Cabral et al., 2006]. Similarly, different reports based on samples from Argentina and Brazil have shown that most of the samples characterized previously as subtype F in one genomic region were indeed BF recombinants in the complete genome, suggesting that the occurrence of pure subtype F is rare [Carr et al., 2001; Guimarães et al., 2002; Thomson et al., 2002; Quarleri et al., 2004; Barreto et al., 2006]. If these mosaic viruses present evolutionary advantages over pure subtype F strains is a subject for future investigation. In this regard, out of the 23 recombinant viruses identified in this work and among five different patterns, 19 had subtype B sequence in env, which could be an indication of selective pressure to the maintenance of the subtype B envelope. Supporting this theory, one previous study [De Sa Filho et al., 2005] showed that among six different BF recombination patterns, all of them were subtype B in the envelope and another one [Monteiro et al., 2007] reported that from three different BC mosaic patterns, recombination always occurred with subtype B in env. On the contrary, these strains could be representing a founder effect event, deriving from a common recombinant ancestral that went through different evolution process and subsequent recombination.

The epidemiological data of the individuals enrolled in this analysis revealed that while subtype B epidemic follows the national tendency [Boletim Epidemiológico AIDS, 2007] for the general population with a male/female ratio of 2:1, the subtype F and F recombinants showed the opposite (sex ratio = 1:1.8) with 68% (19 against 9) of the infected patients being women (Table I). Also, almost 80% were infected heterosexually. These observations suggest an association between subtype F infection and heterosexual transmission (OR = 5.4; 95% CI 0.8756–34.6853) as reported previously [Rios et al., 2005]. The higher prevalence of women in the subtype F epidemic compared to subtype B epidemic coincides with the later introduction of the former in Brazil at the same time that the heterosexual transmission became more frequent. In contrast, the subtype B epidemic started in Brazil when the most common route of transmission of HIV was the homosexual contact followed by the intravenous drug users.

Five different patterns of subtypes B and F recombination between gag and env were identified, suggesting that HIV-1 inter-subtype recombination is occurring very frequently in this area. Among these recombinants, 18 samples shared the same breakpoint in gag. However, they presented different genotypes in env, meaning that (1) they can be the result of different recombination patterns or (2) they shared a common ancestry at some point. Other reports [Carr et al., 2001; Thomson et al., 2002; Quarleri et al., 2004; Sanabani et al., 2006] have described the co-circulation of BF recombinants with different genetic structures but presenting breakpoints at the same or at a very close position in the genome. This suggests the occurrence of hot spots for recombination in the HIV genome. In fact,
the initial region of gag seems to represent one of these hot spots since besides the recombinants identified in this cohort (position 1398 relative to HXB2), other unique BF recombinants [Carr et al., 2001; Thomson et al., 2002; De Sa Filho et al., 2005; Sanabani et al., 2006] as well as the CRF12 from Argentina (position 951) [Thomson et al., 2002], CRF28 and CRF29 (position 1322) [De Sa Filho et al., 2006] from Brazil presented crossover around these genome portions.

Two of the BF recombinants described previously (01BR226 and 02BR33) [Sanabani et al., 2006] grouped together with 11 of the new recombinants in both gag and env phylogenies (Fig. 3). This suggests that they shared common ancestries in both subtypes but as they present different breakpoints, they derived from distinct process of recombination. The same happened for two recombinants from Bahia (155 and 414), which fell in the same group of CRF28 and CRF 29 sequences. Eleven recombinants of this cohort sharing the same breakpoint in gag clustered in a monophyletic cluster apart from the CRF28/29 group pointing to different recombination origins. All recombinants of this group that have available env sequences, but one, also fell into a monophyletic group in both env tree and gag/env concatenated tree. These sequences were unrelated to CRF12. Taken together, these observations represent evidences that a new CRF is circulating in Bahia with a high genetic variability degree, with subtypes B, F, and D and their related recombinant forms being co-circulating. Subtype F was strongly related to heterosexually infected females. Similar to what has been observed for Brazilian southeastern region, variants with different recombination patterns between subtypes B and F, including a potentially new CRF, are present in the northeastern. An increased prevalence of these variants was detected in this geographic area. As

the genetic and antigenic variability of HIV-1 represent a major challenge for the development of globally effective HIV vaccine, the understanding of the mechanisms driving these processes is crucial.

ACKNOWLEDGMENTS

We thank Mrs. Elisabeth Deliege Vasconcelos and Mr. Patrick Younan for editing and revising this manuscript, respectively.

REFERENCES


