Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences

Michael M. Thomson, Elena Delgado, Isabel Herrero, María Luisa Villahermosa, Elena Vázquez-de Parga, María Teresa Cuevas, Rocío Carmona, Leandro Medrano, Lucía Pérez-Álvarez, Laureano Cuevas and Rafael Nájera

Department of Viral Pathogenesis, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220 Majadahonda, Madrid, Spain

The findings that BF intersubtype recombinant human immunodeficiency type 1 viruses (HIV-1) with coincident breakpoints in pol are circulating widely in Argentina and that non-recombinant F subtype viruses have failed to be detected in this country were reported recently. To analyse the mosaic structures of these viruses and to determine their phylogenetic relationship, near full-length proviral genomes of eight of these recombinant viruses were amplified by PCR and sequenced. Intersubtype breakpoints were analysed by bootscanning and examining the signature nucleotides. Phylogenetic relationships were determined with neighbour-joining trees. Five viruses, each with predominantly subtype F genomes, exhibited mosaic structures that were highly similar. Two intersubtype breakpoints were shared by all viruses and seven by the majority. Of the consensus breakpoints, all nine were present in two viruses, which exhibited identical recombinant structures, and four to eight breakpoints were present in the remaining viruses. Phylogenetic analysis of partial sequences supported both a common ancestry, at least in part of their genomes, for all recombinant viruses and the phylogenetic relationship of F subtype segments with F subtype viruses from Brazil. A common ancestry of the recombinants was supported also by the presence of shared signature amino acids and nucleotides, either unreported or highly unusual in F and B subtype viruses. These results indicate that HIV-1 BF recombinant viruses with diverse mosaic structures, including a circulating recombinant form (which are widespread in Argentina) derive from a common recombinant ancestor and that F subtype segments of these recombinants are related phylogenetically to the F subtype viruses from Brazil.

Introduction

Recombination is one of the major mechanisms contributing to retrovirus variability (Katz & Skalka, 1990; Temin, 1991), allowing for the rapid generation of virus variants with increased replicative capacity (Wooley et al., 1997; Gundlach et al., 2000), drug resistance (Gu et al., 1995; Kellam & Larder, 1995; Moutouh et al., 1996) or modified expression of antigenic epitopes (Tumas et al., 1993). Recombination in retroviruses occurs during reverse transcription by the alternate switching of templates between both genomic RNA strands copackaged in each virion, according to the copy-choice model (Coffin, 1996). The importance of recombination as a major factor that contributes to shaping the diversity of human immunodeficiency virus type 1 (HIV-1) genetic forms circulating in the global pandemic was not appreciated until relatively recently (Gao et al., 1996a; Carr et al., 1996; McCutchan et al., 1996). Currently, 11 intersubtype circulating recombinant forms (CRFs) have been reported (Theoretical Biology and Biophysics Group, 2001), although only CRF01_AE of Southeast Asia, CRF02_AG of West Africa, CRF03_AB of Kaliningrad, Russia and CRF07_BC and...
We reported recently the finding of BF intersubtype recombinant viruses that are widely circulating in Argentina (Thomson et al., 2000). This country has, only surpassed by Brazil, the second most numerous population of HIV-1-infected individuals in South America, with an estimated number of 130,000 HIV-1-infections at the end of 1999 (UNAIDS, 2001), which were mainly concentrated in Buenos Aires city and province, where 75% of AIDS cases have been notified (Ministry of Health, 2001). Early in the epidemic, most AIDS cases were diagnosed in homosexual men and subsequently in injecting drug users (IDUs), but, more recently, there has been an increase in infections transmitted heterosexually, particularly among women (Ministry of Health, 2001; Cahn et al., 1998; Vila-Pérez & Bianco, 1998). In our previous study, we detected BF recombinant viruses in 21 (40%), subtype B viruses in 31 (60%) and non-recombinant F subtype viruses in none of 52 samples collected in Buenos Aires between 1995 and 1998 (Thomson et al., 2000). Recombinant viruses were predominant among IDUs and heterosexually infected women, whereas subtype B viruses were predominant among hetero- and homosexually infected men. Coincident breakpoints in pol suggested a common ancestry of the recombinants. These results apparently contradicted other studies (Marquina et al., 1996; Campodonico et al., 1996; Fernández-Medina et al., 1999; Masciotra et al., 2000), suggesting the frequent finding of F subtype viruses (up to 40%) and the relative scarcity (less than 5%) of BF recombinant viruses in Argentina. This apparent discrepancy can be explained by the fact that the segments examined in these studies were, in most recombinants from Argentina, of subtype F, while our analysis included a segment of pol containing intersubtype breakpoints in all of these viruses. Consequently, our results implied that most, if not all, viruses from Argentina, identified previously as being subtype F, were probably BF recombinants. In the 93 samples collected from Buenos Aires in 1999, we have confirmed the absence of non-recombinant F subtype viruses and the high prevalence (65% of samples) of recombinant BF subtype viruses (unpublished data).

For a more complete genetic characterization of recombinant BF viruses from Argentina, we have analysed the near full-length sequences of eight of these viruses, examining recombination points, phylogenetic relationships and the presence of characteristic amino acids and nucleotides. The results indicate that, while there is a considerable diversity of mosaic structures, all recombinant viruses examined appear to share a common ancestry.

### Methods

**Subjects.** The eight subjects studied, five women and three men, attended hospitals in Buenos Aires and had been identified previously, by analysis of partial pol sequences, as harbouring recombinant BF viruses. Three of these subjects were reported previously by Thomson et al. (2000). Of the eight subjects, four subjects (all women) were infected through heterosexual contact and three subjects were IDUs. The risk exposure of one subject was not available. No mutual epidemiological links were known. Samples were collected in 1997 or 1999. Epidemiological data of all subjects are shown in Table 1.

**Sample preparation, PCR amplification and sequencing.** Peripheral blood mononuclear cells were separated by centrifugation on Ficoll–hypaque gradients. Samples were prepared for PCR by cell lysis and digestion with proteinase K, as described previously (Tenorio et al., 1993). A lysate of 1 × 10^6 cells was used for each PCR. Amplification of the near full-length proviral genome (approximately 9 kb) was carried out by nested PCR in four overlapping segments of 1.8–3.2 kb each. Reagents and thermocycling profiles were identical for both first- and second-round PCR amplifications. Each reaction included 2.5 U Taq DNA polymerase, 0.3 µM DNA polymerase, 0.2 mM each dNTP, 0.4 mM each primer, 2 mM MgCl_2, 16 mM (NH_4)_2SO_4, 67 mM Tris–HCl (pH 8.8) and 0.01% Tween-20 in a volume of 50 µl. The thermocycling profile was as follows: initial denaturation for 3 min at 94 °C, 35 cycles of 94 °C for 3 s, 57 °C for 30 s and a final extension at 72 °C for 7 min. For second-round PCR, a 2 µl aliquot of the first-round PCR mixture was used. Sequences of primers are available upon request. Amplification was checked by agarose gel electrophoresis and ethidium bromide staining. After enzymatic removal of the primers and dNTPs that remained in solution (Werle et al., 1994), purified PCR products were sequenced directly in overlapping segments of approximately 500 nt by primer walking using the ABI Prism BigDye Terminator Cycle Sequencing kit and the ABI 377 Sequencer kit (Applied Biosystems). Sequences were corrected and assembled using the BioEdit program (Tom Hall, http://www.mbio.ncsu.edu/BioEdit/bioedit.html). To exclude the possibility of PCR-mediated artefacts, breakpoints were confirmed in duplicate PCR amplifications carried out separately.

**Phylogenetic analysis.** Sequences were aligned using CLUSTAL X (Thompson et al., 1997) with minor manual adjustments considering protein sequences. Phylogenetic neighbour-joining trees (Saitou & Nei, 1987) were based on Kimura’s two-parameter distance matrices (Kimura, 1980) with assessment of the consistency of tree topologies by bootstrapping (Felsenstein, 1985); trees were constructed with CLUSTAL X and viewed with TreeView (Rod Page, http://taxonomy.zoology.gla.ac.uk/rod/treview.html). Analysis of recombination points was done by bootscanning (Salminen et al., 1995) using the Simplot software, version 2.5 (Stuart Ray, http://www.med.jhu.edu/)

### Table 1. Epidemiological data of study subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Age</th>
<th>Risk category</th>
<th>Year of diagnosis</th>
<th>Year of sample collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>A32878</td>
<td>F</td>
<td>31</td>
<td>Heterosexual</td>
<td>1995</td>
<td>1997</td>
</tr>
<tr>
<td>A32879</td>
<td>M</td>
<td>37</td>
<td>IDU</td>
<td>1990</td>
<td>1997</td>
</tr>
<tr>
<td>A32989</td>
<td>M</td>
<td>32</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>A025</td>
<td>F</td>
<td>28</td>
<td>Heterosexual</td>
<td>1991</td>
<td>1999</td>
</tr>
<tr>
<td>A027</td>
<td>M</td>
<td>33</td>
<td>IDU</td>
<td>1985</td>
<td>1999</td>
</tr>
<tr>
<td>A047</td>
<td>F</td>
<td>32</td>
<td>IDU</td>
<td>1988</td>
<td>1999</td>
</tr>
<tr>
<td>A050</td>
<td>F</td>
<td>26</td>
<td>Heterosexual</td>
<td>1999</td>
<td>1999</td>
</tr>
<tr>
<td>A063</td>
<td>F</td>
<td>30</td>
<td>Heterosexual</td>
<td>1996</td>
<td>1999</td>
</tr>
</tbody>
</table>

NA, Not available.
HIV-1 BF recombinants from Argentina

Fig. 1. Bootscan analysis of full-length sequences of two recombinant BF viruses from Argentina. The horizontal axis represents nucleotide distance of the midpoint of the window from the 5’ end of the query sequence (nt 657 in HXB-2). The vertical axis represents the percentage of trees (using 100 bootstrap replicates) that support branching with the consensus subtype reference sequence. A 300 nt window advanced in 20 nt increments was used. Sequences were gap-stripped, transversion to transition ratio was set to 2–0, distances were calculated according to Kimura’s two-parameter model and trees were constructed with the neighbour-joining algorithm.

Results

Recombinant structures

Bootscan plots of the full-length sequences of two viruses are shown in Fig. 1 and the mosaic structures of all eight subtype sequences are available in the Los Alamos Database, a nucleotide had to be absent in all four sequences in order for the signature to be considered subtype B.
Fig. 2. Schematic representation of the mosaic structures of full-length genomes of eight recombinant BF viruses from Argentina. F subtype segments are shown in green and B subtype segments are shown in red. Segments are numbered 1–10 in the 5' to 3' order. LTR sequences that were not analysed are shown in white. Positions in the HXB-2 genome of consensus breakpoints, shown as vertical lines and indicated below, represent midpoints of transition segments between signature nucleotides of different subtypes. A ruler with nucleotide positions in HXB-2 is placed at the bottom.
recombinant viruses, based on bootstrap analysis and inspection of signature nucleotides, are depicted in Fig. 2. Six different mosaic patterns were identified, with multiple intersubtype breakpoints in each virus. Two intersubtype breakpoints were coincident in all and seven in the majority of sequences. All nine ‘consensus’ breakpoints were identified in two viruses (A32789 and A32989) and four to eight were found in each of the remainder six viruses. Positions (HXB-2 numeration) of consensus breakpoints representing midpoints of intersubtype signature nucleotide transitions are indicated at the bottom of Fig. 2. In A32789 and A32989, representative of the consensus mosaic structure (i.e. delimited by consensus breakpoints), there are 10 segments (designated 1–10 in 5’ to 3’ order) of alternating B and F subtypes. In both viruses, subtype B segments comprising approximately 1.6 kb of a predominantly subtype F genome are distributed along the genome as follows: (segment 1) in the 5’ leader sequence and the 5’ segment of p17env; (segment 3) across the protease–reverse transcriptase (RT) border; (segment 5) in the polymerase (p51) domain of RT (this is the longest subtype B segment, approximately 0.7 kb); (segment 7) in the overlap of the first coding exons of tat and rev and in the 5’ segment of vpu; and (segment 9) in the overlap of gp41 and the second coding exon of rev. In segment 3, the bootstrap value that supports the grouping of the recombinant viruses with subtype B did not reach significant values, probably due to its short length and high sequence conservation. This resulted in few phylogenetically informative sites, although inspection of signature nucleotides revealed the presence of four subtype B and no subtype F signatures, thus supporting its tentative assignment to subtype B. A027, A047 and A050 exhibited mosaic structures highly similar to A32789 and A32989, but A027 and A047 lacked a B subtype segment in p17env and A050 had a short, extra B subtype segment near the 5’ end of nef.

A32878, A025 and A063, while sharing several breakpoints with the other five viruses, had different mosaic structures in substantial portions of their genomes, containing additional subtype B segments that were absent from the other viruses. In A32878, most of pol (including integrase, all of the RNase H domain and most of the polymerase domain of RT), and the entire first coding exon of tat are subtype B segments. In A025, subtype B segments include the entire p17env protein, the 5' end of p24ag, most of pol (except protease and the 3' end of integrase) and a segment in the non-coding 3' U3 region. In A063, B subtype segments comprise most of pol, the 3' half of vif, most of env and a segment near the 5' end of nef, the last coinciding with the B subtype segment at an identical site in A050. A063 is the only virus with a genome that is predominantly subtype B.

Phylogenetic analysis of partial segments

Phylogenetic neighbour-joining trees of partial segments delimited by consensus breakpoints (i.e. those present in the majority of viruses and coinciding with the breakpoints of A32789 and A32989) (Fig. 3) confirm the subtype assignments of segments obtained by bootscanning. In the bootscanning analyses, bootstrap support of segment 3 clustering with subtype B reference viruses did not reach significant values (data not shown). The phylogenetic trees provide additional phylogenetic information: (i) subtype F segments 6 and 8 (excluding, in each tree, viruses that contain breakpoints in the corresponding segment) cluster with F subtype viruses from Brazil, with highly significant bootstrap values (Fig. 3c, d), suggesting a Brazilian ancestry of these segments; (ii) a tree of concatenated sequences obtained by joining all subtype F segments of each virus also supports the clustering of A32789, A32989, A027, A047 and A050 with each other, with 91% bootstrap value, and with the F subtype reference virus 93BR020 from Brazil, with 100% bootstrap support (Fig. 3f); (iii) in subtype B segment 5, located in RT, all recombinant BF viruses from Argentina, except A32878, cluster together, apart from five subtype B viruses from Argentina, with 63% bootstrap value, which increases to 78% when A32788 is excluded; shorter branches of recombinants relative to B subtype viruses from Argentina are consistent with a more recent common ancestry of the former (Fig. 3b); (iv) a tree of a B subtype segment in pol found only in A32788, A025 and A063 supports the phylogenetic relationship of A32788 and A063 with the BF recombinant virus 93BR029 from Brazil, but A025 branches separately (Fig. 3e). Branching of A32788 in segment 5 apart from the other viruses suggests that the B subtype sequences in pol of this virus derive from a second B subtype virus that is unrelated to the parent of segment 5 in the other viruses; the fact that A063 is phylogenetically related to A32788 in the 3' half of pol suggests that this segment of A063, and possibly also the B subtype segments in vif and env that are unique to this virus, have an origin different from the parental segments of the other B subtype viruses that are common to all recombinants.

Phylogenetic tree of full-length sequences

In the phylogenetic tree of the full-length genome (Fig. 4), the five viruses exhibiting similar mosaic structures, A32789, A32989, A027, A047 and A050, group in a monophyletic cluster supported by 100% bootstrap value. All viruses except A063, which clusters with subtype B sequences, branch with subtype F reference isolates, reflecting the predominance of subtype F along their genomes. When the B subtype segments of A32788, A025 and A063, found only in these viruses, are excluded, clustering of the remaining portion of their genomes with the five other isolates is supported by 100% bootstrap values (data not shown).

Recombinant structures and protein functional domains

The functional domains of the BF chimeric proteins p17env, RT, Tat, Rev, Vpu and gp41 of the consensus recombinant
Fig. 3. Neighbour-joining phylogenetic trees of partial segments of recombinant BF viruses from Argentina. Viruses with intersubtype breakpoints in the segments analysed were excluded from each analysis of partial sequences. Nucleotide positions (HXB-2 numeration) that delimit the analysed fragments and the segment number according to Fig. 2 are shown above each tree. Trees are rooted with simian immunodeficiency virus isolate cpzUS. Recombinant BF viruses from Argentina are shown in boldface and F subtype viruses from Brazil are underlined. B subtype viruses from Argentina, sequenced in full-length or partial pol sequences, are marked with an asterisk and subtype reference sequences are indicated with the subtype designation followed by the name of the isolate. Bootstrap values $\geq 70\%$ (based on 100 replicates) are shown. The bootstrap value shown in parentheses in (b) was obtained after excluding A32878 from the analysis. Relevant clusters including BF recombinants from Argentina and supported by $\geq 70\%$ bootstrap values are indicated in square brackets.
HIV-1 BF recombinants from Argentina

Fig. 4. Unrooted neighbour-joining phylogenetic tree of full-length genomic sequences of recombinant BF viruses from Argentina. Recombinant viruses are shown in boldface and subtype reference viruses are indicated by the subtype designation followed by the name of the isolate. Bootstrap values > 70%, based on 100 replicates, of some key nodes are shown.

Characteristic amino acids and nucleotides

Examination of the predicted amino acid sequences of viral proteins revealed some residues that are highly characteristic of the BF recombinants from Argentina. In Table 2, residues found in three or more recombinants, unreported in F1 subtype viruses and either unusual or unreported in B subtype viruses in the Los Alamos Database (Theoretical Biology and Biophysics Group, 2001) are shown. Of these, three are highly unusual in all group M viruses. Vif A^41, identified in four of eight BF recombinants from Argentina, is absent from all 686 group M Vif sequences in the Los Alamos Database alignments. Moreover, a non-acidic amino acid at position 61 of Vif is highly unusual (present in 1.8%) in group M viruses, whereas in the BF recombinants from Argentina, only one of eight viruses has an acidic amino acid at this position. Tat N^65, present in all eight BF recombinant BF viruses from Argentina, is found in only 24 (3.1%) of 771 group M Tat sequences in the Los Alamos Database alignments, including 17 subtype B sequences derived from a single individual. gp120 S^412, identified in three BF recombinant viruses from Argentina, is unreported among B or F1/F2 subtype viruses and is present in only 11 (1.5%) of 740 group M Env sequences in the Los Alamos Database alignments.

Several synonymous nucleotide substitutions in coding regions and other nucleotides in non-coding regions highly characteristic of the BF recombinants from Argentina were also identified. Of these nucleotides, 12 were present in ≥ 50% of the recombinants from Argentina and absent from all six F1/F2 subtype viruses and not found or highly unusual (≥ 2 of 64 full-length sequences) in subtype B references: A^929, T^950, G^1056, G^1118, T^1227, T^1220, C^1252, T^1466, T^1411, C^846, C^926 and T^9419 (positions are numbered according to the HXB-2 genome). Each recombinant virus from Argentina has 5–11 of these characteristic nucleotides. Among these, two nucleotides are either highly unusual or unreported in HIV-1 isolates: C^9456 in the 3' U3 region, found in seven of eight BF recombinants from Argentina, is present in only 2 (1%) of 200 full-length HIV-1 sequences in the Los Alamos Database alignments; and T^9419, at position 7 of the decanucleotide SP1-II-binding site of the 3' U3 region (AGGGCGTGAC) found in five BF recombinants from Argentina, is absent from all 587 HIV-1 sequences spanning this segment in the Los Alamos Database alignments (in which a G is present in all cases except one). In the SP1-III site, the reverse substitution occurs: a pyrimidine (usually a T) at position 7 (HXB-2 position 9468), which is highly conserved among all group M subtypes, is replaced by
Fig. 5. Recombinant structures and functional domains of seven proteins of prototypical recombinant BF viruses from Argentina. B subtype segments are shown in red and F subtype segments are shown in green. Subtype segments are delimited by the midpoints of intersubtype nucleotide signature transitions.

a G in seven of the BF recombinants from Argentina (except A025, which is subtype B in this segment).

The crown tetrapeptide of the Env V3 loop is GPGR in five recombinants, GPGQ in two and GWGR in one. GPGR is frequent in F subtype viruses from Brazil and uncommon in other F subtype isolates. GWGR is present in A063, the only virus with gp120 of subtype B. This motif is characteristic of B subtype viruses from Brazil, found in 40% of these viruses (Potts et al., 1993; Morgado et al., 1994, 1998) and uncommon elsewhere. In phylogenetic trees of gp120, A063 clustered
Table 2. Signature amino acids of recombinant BF viruses from Argentina

Amino acid positions (in parentheses) are numbered according to the proteins of the HIV-1 NL4.3 isolate. Signature amino acids of BF recombinants from Argentina are shown in boldface.

<table>
<thead>
<tr>
<th></th>
<th>p2gag (9)</th>
<th>p6gag (13)</th>
<th>Protease (93)</th>
<th>RT (281)</th>
<th>RT (480)</th>
<th>Integrase (122)</th>
<th>Vif (61)</th>
<th>Tat (65)</th>
<th>Vpu (75)</th>
<th>Vpu (77)</th>
<th>gp120 (442)</th>
<th>gp41 (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B subtype consensus</td>
<td>I</td>
<td>Q</td>
<td>I</td>
<td>K</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>H</td>
<td>P</td>
<td>D</td>
<td>Q</td>
<td>I</td>
</tr>
<tr>
<td>F1 subtype consensus</td>
<td>I</td>
<td>R</td>
<td>L</td>
<td>G</td>
<td>H</td>
<td>P</td>
<td>A</td>
<td>N</td>
<td>L</td>
<td>N</td>
<td>I/L</td>
<td></td>
</tr>
<tr>
<td>A32879</td>
<td>V</td>
<td>W</td>
<td>L</td>
<td>G</td>
<td>H</td>
<td>P</td>
<td>A</td>
<td>N</td>
<td>L</td>
<td>N</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>A32989</td>
<td>V</td>
<td>W</td>
<td>L</td>
<td>R</td>
<td>H</td>
<td>P</td>
<td>A</td>
<td>N</td>
<td>L</td>
<td>N</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>A047</td>
<td>V</td>
<td>W</td>
<td>L</td>
<td>R</td>
<td>Q</td>
<td>T</td>
<td>T</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>A027</td>
<td>V</td>
<td>W</td>
<td>L</td>
<td>R</td>
<td>Q</td>
<td>T</td>
<td>T</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>S</td>
<td>I</td>
</tr>
<tr>
<td>A050</td>
<td>V</td>
<td>W</td>
<td>L</td>
<td>R</td>
<td>H</td>
<td>P</td>
<td>A</td>
<td>N</td>
<td>L</td>
<td>N</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>A32878</td>
<td>V</td>
<td>R</td>
<td>L</td>
<td>K</td>
<td>Q</td>
<td>S</td>
<td>T</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>S</td>
<td>I</td>
</tr>
<tr>
<td>A025</td>
<td>V</td>
<td>R</td>
<td>L</td>
<td>K</td>
<td>Q</td>
<td>S</td>
<td>T</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>S</td>
<td>I</td>
</tr>
<tr>
<td>A063</td>
<td>V</td>
<td>R</td>
<td>L</td>
<td>K</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>N</td>
<td>P</td>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>BF recombinants*</td>
<td>100</td>
<td>62:5</td>
<td>62:5</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>62:5</td>
<td>100</td>
<td>37:5</td>
<td>62:5</td>
<td></td>
</tr>
<tr>
<td>B subtype*</td>
<td>8:5</td>
<td>0</td>
<td>4:5</td>
<td>0</td>
<td>4:5</td>
<td>9</td>
<td>0</td>
<td>4:5</td>
<td>0</td>
<td>4:5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F1 subtype*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Percentage of viruses in which amino acids shown in boldface are found in BF recombinants from Argentina (this study) and in B or F1 subtype viruses in the Los Alamos HIV Sequence Database alignments (Theoretical Biology & Biophysics Group, 2001).

with a group of subtype B viruses from Brazil, including all those containing the GWGR crown motif and excluding those with the GPGR tetrapeptide, although bootstrap values (57%) did not reach significance (data not shown).

Discussion

The results of this and our previous study on recombinant BF subtype viruses from Argentina underscore the importance of analysing full-length genome sequences to appreciate more completely the diversity of genetic forms circulating in a geographical area, particularly where diverse genetic forms of HIV-1 are cocirculating. In none of four studies in Argentina by other authors, where each study was based on the analysis of short sequences, was the real extent of the circulation of HIV-1 BF recombinant viruses recognized, since the segments of short sequences, was the real extent of the circulation of HIV-1 are cocirculating. In none of four studies in Argentina in any geographic area, particularly where diverse genetic forms of HIV-1 are cocirculating. In none of four studies in Argentina by other authors, where each study was based on the analysis of short sequences, was the real extent of the circulation of HIV-1 BF recombinant viruses recognized, since the segments analyses were, in most recombinants, subtype F (Marquina et al., 1996; Campodonico et al., 1996; Fernández-Medina et al., 1999; Masciotra et al., 2000). Only after the analysis of a segment in pol containing breakpoints in all recombinant viruses from Argentina was it realized that BF subtype recombinant viruses were circulating widely and that non-recombinant F subtype viruses could not be detected (Thomson et al., 2000); this implies that viruses classified previously as being subtype F viruses were probably recombinant BF subtype viruses. In the present study, we extend this novel finding by analysing the near full-length genome of eight recombinant BF subtype viruses from Argentina. Examination of the mosaic structures (Fig. 2) revealed six different patterns, although in a group of five viruses, with a predominantly F subtype genome, differences with each other were minor. Nine consensus breakpoints were identified: two of them were found in all eight viruses and seven in the majority. Two recombinants, A32879 and A32989, with coincident mosaic structures contained all nine consensus breakpoints, which delimit 10 segments of alternating B and F subtypes (designated 1–10). These viruses will be referred to as the prototypical BF recombinant viruses from Argentina. Partial pol and gag sequences of additional recombinant viruses from Argentina, not included in this study (Thomson et al., 2000; unpublished data), show mosaic structures that are, in most of them, coincident with those of A32879 and A32989. Sharing of several cross-over sites by all or the majority of isolates of this study suggests that they derive from a common recombinant ancestor, with subsequent recombination resulting in different mosaic patterns, as in the case of A32878, A063 and A025.

Further support for a common ancestry was obtained by phylogenetic analysis of partial segments delimited by breakpoints (Fig. 3), which also suggested a Brazilian ancestry of the F subtype segments. In the longest B subtype fragment common to all viruses (located in RT), all recombinants except A32878 formed a cluster separate from the F subtype viruses from Argentina (Fig. 3b). It is most parsimonious to assume that the remainder B subtype segments shared by all viruses probably derive from the same parental virus, considering the coincident breakpoints delimiting these segments. Clustering with F subtype viruses from Brazil was observed in two of the F subtype segments (Fig. 3c, d), as well as in concatenated sequences obtained by joining all F subtype segments of the
recombinants (Fig. 3f); in the last tree, the five viruses exhibiting similar mosaic structures formed a cluster supported by a 91% bootstrap value (Fig. 3f). The common ancestry of these five viruses was supported strongly in phylogenetic trees of full-length genomes (Fig. 4). Furthermore, when B subtype segments, found only in the remaining three viruses that showed divergent mosaic structures, were excluded from the analysis, a common origin of each virus with the other five isolates was also supported strongly in phylogenetic trees (data not shown). In two viruses, A32878 and A063, some B subtype segments appeared to derive from a virus unrelated to the parental virus of the prototypical recombinants. Clustering of these segments with the BF recombinant isolate 93BR029 from Brazil (Fig. 4e) and the presence in one virus of the GWGR V3 crown tetrapeptide, characteristically found in 40% of B subtype viruses from Brazil (Potts et al., 1993; Morgado et al., 1994, 1998) and which is uncommon in Argentina (identified in only 1 of 24 B subtype V3 sequences from Argentina in the Los Alamos Database), suggested a Brazilian ancestry of the extra B subtype segments.

The presence of highly characteristic amino acid residues (Table 2) and nucleotides shared by all or the majority of the BF recombinants from Argentina also argues in favour of a common ancestry of these viruses. Two of these amino acids, Tat N65 and Vpu N77, were found in all of the recombinants. Notably, Vif A41, found in four viruses, is absent from all 86 HIV-1 group M sequences in the Los Alamos Database and Tat N65 is found in only 1·8% of group M sequences in the Database. Of the nucleotide substitutions that do not involve amino acid changes, the presence of a G to T mutation at position 7 of the decanucleotide SP1-II-binding site in the 3' U3 region, found in five BF recombinants, is remarkable, as it is found in none of the 587 HIV-1 sequences in the Los Alamos Database. Similar to the SP1-II site of recombinants from Argentina, the SP1-I and SP1-III sites of other HIV-1 isolates usually have a T at position 7 instead of the consensus G of the SP1-binding sites and this substitution, in the context of the HIV-1 long terminal repeat (LTR), does not appear to diminish in vitro binding of the SP1 transcription factor (Jones et al., 1986).

Seven proteins, matrix, protease, RT, Tat, Rev, Vpu and gp41, exhibit chimeric structures in most recombinant viruses from Argentina. The spatial correlation of subtype segments with functional domains is shown in Fig. 5. Interestingly, the short B subtype segment in the cytoplasmic tail of gp41 matches with α-helix 2, which has been proposed on the basis of mutational studies, to interact with residues in the N-terminal of the matrix protein (Murakami & Freed, 2000; Cosson, 1996), which is also subtype B in six recombinants, for the incorporation of Env to the virions. Similarly, the F subtype of Rev, present in all viruses except A063, roughly coincides with the basic domain that interacts with the Rev responsive element (RRE) (Pollard & Malim, 1998), which is also subtype F in these viruses. In A063, both Rev (except a short segment comprising eight amino acids) and RRE are subtype B. Whether subtype coincidence of Rev–RRE and gp41–matrix interacting surfaces in the BF recombinants from Argentina is structurally favourable for increased affinity and beneficial for virus fitness remains to be determined.

The introduction of BF recombinant viruses in Argentina does not seem to be recent. Three IDUs harbouring recombinant viruses studied by us are known to have been infected by 1985 and one heterosexually infected man was infected by 1986. Masciofra et al. (2000) report a case of an IDU infected with Fen subtype virus (probably a BF recombinant) who first tested HIV-1-positive in 1987. Such an early introduction of the recombinant BF viruses in Argentina is consistent with mean intersubject genetic distances in the env V3 region (Thomson et al., 2000). Consequently, the recombinant BF viruses from Argentina might be the earliest known HIV-1 circulating intersubtype recombinant viruses to have originated outside Africa. A number of reasons point to a probable initial introduction of BF recombinants in Argentina among IDUs: (i) most HIV-1-infected IDUs harbour recombinant BF viruses [25 of 31 (81%) samples studied]; (ii) most cases of recombinant BF virus infection with earlier dates of HIV-1 diagnosis are IDUs; (iii) the HIV-1 epidemic among homosexually infected women, the other group in which BF recombinants are predominant, is relatively recent: up to 1990, only 2% of the AIDS cases reported were women infected sexually, as compared to 31% in IDUs (Ministry of Health, 2001); and (iv) all epidemics with intersubtype recombinant forms of non-African origin, CRF03_AB (Liitsola et al., 1998), CRF07_BC (Su et al., 2000), CRF08_BC (Piyasirisilp et al., 2000) and recombinant BG viruses of Spain (Thomson et al., 2001; unpublished data that show a newly characterized CRF), have been identified among IDUs.

Whether the ancestor of the recombinant viruses from Argentina originated locally or in Brazil is not known, but the absence of non-recombinant F subtype viruses in Argentina in 145 samples analysed by us suggests a probable Brazilian provenance of the recombinants. However, recombinant viruses related to those from Argentina appear to be either absent or not circulating widely in Brazil, since none of the BF recombinant sequences from Brazil reported to date (Sabino et al., 1994; Gao et al., 1996b, 1998; Morgado et al., 1994; Cornelissen et al., 1997; Brindeiro et al., 1999; Pilcher et al., 1999) nor any of the five recombinant BF viruses of this country analysed by us in gag and pol (unpublished data) exhibit a mosaic structure analogous to the prototypical recombinants from Argentina. An alternative possibility is that the individual source of the ancestor of the recombinants from Argentina resided and transmitted the recombinant form(s) in Argentina after acquiring the F subtype parental virus in Brazil.

The circulation in a geographical area of distinct but related recombinant forms has been reported previously (McCuchan et al., 1999; Janssens et al., 2000; Cornelissen et al., 2000; Motomura et al., 2000), although the degree of diversity of
recombinant forms of a common ancestry identified in Argentina in this study is higher than that found in other areas. Previous reports of recombinant BF viruses from Argentina with mosaic structures that differ in partial segments from those described here (Marquina et al., 1996; Campodonico et al., 1996; Fernández-Medina et al., 1999; Masciotra et al., 2000), as well as our unpublished analysis of partial sequences, indicate that additional recombinant forms may be present in Argentina. Also, various BF recombinant forms have been identified, by us (unpublished data) and other authors, in Brazil by the analysis of partial sequences, although evidence for a common ancestry is lacking. Several factors may contribute to the high diversity of related recombinant forms in Argentina: (i) the relatively long period in which recombinant BF viruses from Argentina appear to have been in circulation, increasing the chances of recombination with other genetic forms; (ii) the cocirculation of recombinant B and BF subtype viruses in the same population; (iii) the high prevalence of HIV-1 infections among IDUs in Argentina [ranging from 20 to 92% in different surveys (UNAIDS, 2001)], increasing the possibility of coinfections by needle sharing; and (iv) the analysis of full-length sequences of a relatively large number of isolates (compared to other studies), favouring the identification of differences in their recombinant structures.

The distribution of genetic forms in Argentina resembles, in some respects, that of Thailand, with a double epidemic: one of subtype B and another of closely related recombinant BF viruses that were probably introduced later, according to relative branch lengths in pol (Fig. 3b), and distributed unevenly among groups with different epidemiological characteristics. And again, similar to Thailand, Argentina has an apparent expansion of recombinants over the years. Of the 93 samples collected in 1999, BF recombinants increased from 50% in individuals diagnosed before 1993 to 72% in those diagnosed since 1993 (unpublished data). A similar temporal trend in the increase in \( F_{\text{env}} \) viruses (probably BF recombinants) was noticed in a previous study (Masciotra et al., 2000). Prospective studies might reveal if differences in properties of transmission contribute to unequal distribution of genetic forms among epidemiological groups and to temporal changes in the prevalence of the recombinant BF viruses from Argentina.

We have found in other countries BF recombinants related to those from Argentina. In Spain, we have identified two individuals, an Argentinian man and a Spanish woman, who was probably infected through sexual contact with a South American man, harbouring recombinant BF viruses related phylogenetically to the recombinant viruses from Argentina (unpublished data). In Venezuela, a woman, infected by her husband who acquired the infection in Argentina, harboured a recombinant BF virus with a breakpoint in pol identical to the A025 isolate reported here, although full-length pol sequences show differences in their mosaic structures (Delgado et al., 2001). Viruses acquired presumably in Argentina and identified in partial sequences as being subtype F viruses from Bolivia (Velarde-Dunois et al., 2000), Peru (Russell et al., 2000) and Spain (Holguin et al., 2000) might also be BF recombinants related to those reported here.

In summary, analysis of full-length HIV-1 genomic sequences reveals a considerable diversity but a common ancestry of recombinant BF viruses from Argentina and the phylogenetic relationship of these viruses with subtype F viruses from Brazil. The majority of recombinants exhibited similar mosaic structures, but some had unique patterns in part of their genomes, suggesting their generation by successive recombination of a common recombinant ancestor. With the identification of two viruses with identical mosaic structure, which was confirmed in several partial sequences (Thomson et al., 2000; unpublished data), the requirements to define an HIV-1 CRF are fulfilled (Robertson et al., 2000) and would, according to the nomenclature currently accepted, be designated CRF12_BF; [the full-length sequences of three HIV-1 isolates from Argentina and Uruguay proposed to represent a circulating BF recombinant form (CRF12_BF) was announced at the Los Alamos HIV Sequence Database website (http://hiv-web.lanl.gov) by other authors in the very recent past]. This is the first CRF identified to have originated from the Americas and, most likely, the oldest of the known CRFs to have originated outside Africa. It is of note that all known CRFs of non-African origin involve a parental subtype B virus, probably because of the circulation of B subtype viruses in the IDU populations among which they apparently originated; this parallels the predominance of subtype A viruses among the recombinant viruses of African origin. It is predictable that the cocirculation of diverse HIV-1 genetic forms that are, increasingly, brought into contact by international travel (Thomson & Nájera, 2001) will result in the identification of additional CRFs outside Africa. Analysis of full-length sequences of additional isolates from Argentina, Brazil and neighbouring countries will provide a more comprehensive picture of the spectrum of HIV-1 genetic forms circulating in South America and will contribute to the understanding of the mechanisms governing the generation of HIV-1 diversity and its impact on the progression and control of the epidemic.

We thank Horacio Salomón and Maria Ávila for providing clinical samples from Argentina, Amílcar Tanuri and Mariza Morgado for providing clinical samples from Brazil, Saladin Osmanov and José Esparza for organizing the UNAIDS program that made this study possible and Francisco Parras for his support of this study. This work was financed by grant 98BVII236 of Plan Nacional del SIDA, Ministerio de Sanidad y Consumo, Spain and by Technical Service Agreement HQ/98/457048, UNAIDS.

References


Morgado, M. G., Guimarães, M. L., Gripp, C. B., Costa, C. I., Neves, I., Jr, Veloso, V. G., Linhares-Carvalho, M. I., Castello-Branco, L. R.,


Theoretical Biology and Biophysics Group (1999). Human retroviruses and AIDS 1999: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.


**HIV-1 BF recombinants from Argentina**