Infection, Genetics and Evolution 12 (2012) 443–447

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Characterization of full-length HIV-1 CRF17_BF genomes and comparison to the prototype CRF12_BF strains ☆

Paula C. Aulicino [a,*], Manuel Gómez-Carrillo [b], Gonzalo Bello [c], Carlos Rocco [a], Andrea Mangano [a], Jean Carr [d], Luisa Sen [a], Brian Foley [e]

[a] Laboratorio de Biología Celular y Retrovirus-CONICET, Hospital de Pediatría J. P. Garrahan, Buenos Aires, Argentina
[b] National Reference Center for AIDS, Department of Microbiology, School of Medicine, University of Buenos Aires, Buenos Aires, Argentina
[c] Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, Brazil
[d] Department of Epidemiology and Prevention, Institute of Human Virology, University of Maryland School of Medicine. Baltimore, MD 21201, USA
[e] Theoretical Biology Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Article history:
Received 21 March 2011
Received in revised form 4 January 2012
Accepted 5 January 2012
Available online 12 January 2012

Keywords:
HIV-1 subtypes
Argentina
BF recombinants
HIV diversity
Molecular epidemiology
CRF17_BF

ABSTRACT

The aim of this work is to characterize the full-length intersubtype recombinant structure of the HIV-1 Circulating Recombinant Form CRF17_BF. A single genome of CRF17_BF was originally described in 2001 as being largely similar to CRF12_BF. Since then, more genomes of CRF17_BF have been sequenced but not adequately described in publications. Here we describe CRF17_BF as a genuine CRF, and analyze its recombination pattern based on bootscan analyses, subtype signature patterns, and phylogenetic reconstruction of subtype-delimited segments. We show that CRF17_BF can be distinguished from CRF12_BF in several regions of the genome, including *vpu*, *pol*, *env* and *nef*. A complete and accurate characterization and description of recombination breakpoints in CRFs is required for a proper surveillance of HIV-1 genotypes, and important for epidemiological purposes.

1. Introduction

Genetic recombination is one of the main sources of HIV-1 diversity. When recombination occurs between two or more subtypes, new inter-subtype recombinant forms emerge. When one of them is transmitted and reaches a detectable level in the population, gaining epidemiological importance, it is defined as a Circulating Recombinant Form (CRF). By definition, CRFs are inter-subtype recombinant HIV-1 strains or lineages that have been identified in three or more people with no direct epidemiologic linkage. Usually, full-length sequencing is required from at least two of them, and partial sequences spanning recombination sites can complete the characterization. HIV-1 intersubtype recombinants which have been detected and characterized in only one or two patients are described as unique recombinant forms (URFs). CRF17_BF was described in 2001 based on the recombinant structure of a single sequence: ARMA038 (GenBank Accession number AY037281) (Carr et al., 2001). More recently, new related full-length HIV-1 genomes with similar genomic structures have been submitted to the GenBank. The aims of this study are to determine the genuine CRF nature of ARMA038 and related sequences, and compare the recombination patterns between CRF17_BF and CRF12_BF, the prototypic BF recombinant HIV-1 strain in South America.

2. Materials and methods

Seven full-length HIV-1 nucleotide sequences from Argentina (AR.1999.ARMA038 [AY037281], AR.2002.AR02-ARG1139 [EU581825], AR.2002.AR02-ARG2233 [EU581826]); Bolivia (BO.2002.BO02-BOL119 [EU581827]); Paraguay (PY.2002.PY02-PSP0096 [EU581823], PY.2002.PY02-PSP0073 [EU581824]); and Peru (PE.2002.PE02-PCR0155 [EU581828]) denoted as CRF17_BF were retrieved from the Los Alamos HIV Sequence Database and aligned with CRF12_BF, B, C, D and F1 subtype reference sequences using the GeneCutter Tool (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). The monophyletic nature of the full-length and concatenated B or F1 segments was evaluated in maximum likelihood (ML) phylogenetic trees constructed with MEGA 5.05 (Tamura et al., 2011) under a General Time Reversible (GTR)

☆ This work was the result of a collaborative effort fueled by the International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology (http://www.rega.kuleuven.be/cev/workshop/).

* Corresponding author. Address: Hospital de Pediatría J. P. Garrahan, Laboratorio de Biología Celular y Retrovirus, Combate de los Pozos 1881 (1245), Ciudad Autónoma de Buenos Aires, Argentina. Tel.: +54 11 4308 1998; fax: +54 11 4308 5325.

E-mail addresses: pauauli@gmail.com, paulicino@garrahan.gov.ar (P.C. Aulicino).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2012.01.003

model of evolution, Gamma distribution of site-specific rates of evolution with eight rate categories. Tree figures were rendered from the MEGA tree outfile using FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Recombination breakpoints were characterized after unifying and confirming the results obtained by different methods:

(i) Similarity and bootscanning plots were performed with Simplot software version 3.5.1 (Lole et al., 1999). In bootscan analyses, bootstrap values were determined in Neighbor-Joining (NJ) trees constructed using the Kimura two-parameter model, based on 100 re-samplings, supporting branching with the consensus sequences within a 200 bp window moving in steps of 20 bases. Subtypes F1 and B were used as references, and subtype C ZA.2004.SK164B1 (AY772699) was used as an outgroup.

(ii) RIP 3.0 program (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) was used to identify BF recombination segments in the CRF17_BF genomes, by comparing it to a set of subtype reference sequences using a 200 bp window, and a 90% confidence level. In RIP, the window is moved in increments of one nucleotide residue from left to right in the alignment, and a Hamming distance (p-distance) is calculated for each window. The best match within each window is qualified by a measure of confidence -which is approximate- obtained by comparing the distance to the first and second best-matching reference sequences by using a *z*-test as statistical support.

(iii) Nucleotide alignments were visually inspected to identify subtype B and F1 signature patterns. Signature nucleotides that discriminate between subtypes B and F1 were defined as those found in at least 90% of the reference sequences of one subtype and in less than 10% of those of the other. Recombination breakpoints were established at the middle of a genomic region flanked by signature nucleotides of the two subtypes.

(iv) NJ phylogenetic trees were built for individual subtype B segments delimited by inter-subtype recombination breakpoints in CRF17_BFs, using the Tamura-Nei evolutionary model incorporated in the MEGA 5.05 program.

Subtype reference sequences used for identification of recombination breakpoints and phylogenetic reconstructions were the following: **CRF12_BF:** UY.1999.URTR23 (AF385934), AR.1997.A32879 (AF408629), AR.1997.A32989 (AF408630), UY.1999.URTR35 (AF385935), AR.1999.ARMA159 (AF385936); **subtype F1:** BR.1990.BZ163 (AY173958), BR.2002.02BR082 (FJ771006), BR.2002.02BR170 (FJ771007), BR.2006.06BR564 (FJ771008), BR.2006.06BR579 (FJ771009), BR.2007.07BR844 (FJ771010), FI.1993.FIN9363 (AF075703), BR.1993.93BR020-1 (AF005494), BE.1993.VI850 (AF077336), AR.2002.ARE933FL (DQ189088.1); **subtype B:** TH.1990.BK132 (AY173951), US.1990.WEAU160_GHOSH (U21135), NL.2000.671-00T36 (AY423387), US.1998.1058-11 (AY331295), FR.1983.HXB2-LAI-IIIB-BRU (K03455), US.1986.JRFL_JR-FL (U63632), AR.ARMA132 (AY037282), AR.ARMA173 (AY037274), AR.ARCH054 (AY037268.1), BO.BOL122 (AY037270.1).

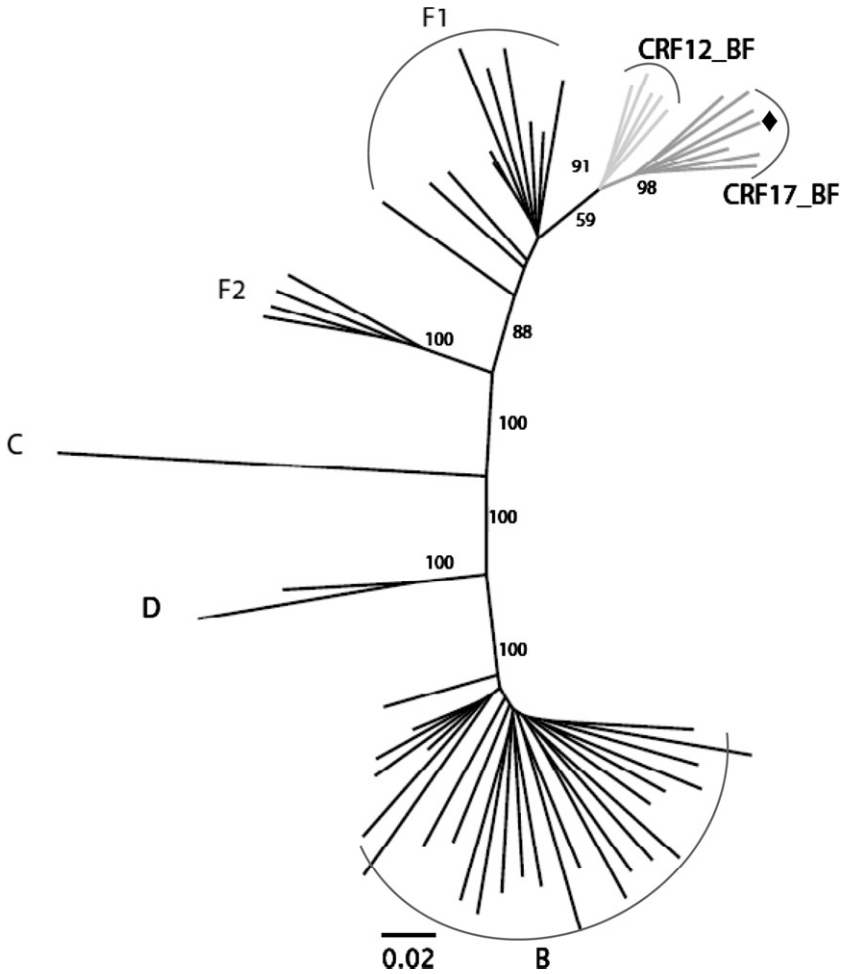

Fig. 1. Monophyletic clustering of full-length CRF17_BF HIV-1 strains. A Maximum likelihood tree was constructed with CRF17_BF-like and subtype reference sequences to show the monophyletic origin of six recently described strains with ARMA038 (♦). Bootstrap values based on 1000 replicates are shown on the nodes.

The multiple sequence alignments used for analyses are available from the authors on request.

3. Results

3.1. Identification of CRF17_BF as a genuine CRF

A Maximum likelihood phylogenetic tree was constructed for full-length HIV-1 reference sequences of the following clades: B, C, D, F1, CRF12_BF and the putative CRF17_BF. All seven genomes previously classified as CRF17_BF clustered together, independently of the CRF12_BF clade, confirming both the monophyletic nature of the CRF17_BF strains and the close but distinct relationship of CRF17_BF with CRF12_BF strains (Fig. 1). Epidemiological linkage was only suspected for sequences PSP0073 and PSP0096 from Paraguay, due to the high similarity between them (data not shown).

3.2. Characterization of the BF recombinant structure in CRF17_BF genomes

In order to identify subtype B and F1 genomic segments on CRF17_BF, we characterized the subtype structure of the seven HIV-1 genomes contained in the CRF17_BF cluster by similarity plots and bootscan analyses with Simplot software version 3.5.1 (Lole et al., 1999). All HIV-1 genomes had the same recombinant structure as the ARMA038 strain on bootscanning plots with a window size of 200 bases, sharing nine subtype B insertions on a subtype F1 background structure (Figs. 2a and b). Each subtype B segment was identified with a letter from A to I, and distributed along the HIV-1 genome as follows: segment A in the 5′ portion of p17*gag*; segment B across the protease-reverse transcriptase (RT) border; segment C in the polymerase (p51) domain of RT; segment D in the overlap of the first coding exons of *tat, rev*, *vpu* and the 5′ segment of *env*; segment E across the gp120-gp41 border of *env*; segment F in the gp41*env*; segment G in the overlap of the second coding exon of *rev* and the 3′ end of gp41; segment H in the 5′ end of *nef*; and segment I in 3′ the overlap of *nef* and the 3′ leader sequence.

Almost the same recombination pattern was obtained after submitting the alignment to RIP 3.0 program, again with a window size of 200 bases. RIP indicated that the similarity to subtype B was not statistically significant at the 95% setting for fragments H and I. The only difference in subtype assignment between bootscanning and RIP analysis was found for fragment F, which was subtype B by bootscanning and subtype F1 by RIP. Further confirmation of the proposed recombination pattern was obtained after visual inspection of subtype signature patterns on the nucleotide sequence alignment, and phylogenetic reconstruction of independent subtype B delimited segments as detailed below. The same analysis was done for subtype F1 delimited segments (data not shown). Segment H was defined mainly based on subtype signature nucleotides. ML phylogenetic trees of concatenated subtype F1 (Fig. 2c) and subtype B (Fig. 2d) segments defined by bootscanning confirm the monophyletic nature of CRF17_BF sequences, supporting the notion that all of these recombinants genomes are derived from a common BF recombinant ancestor.

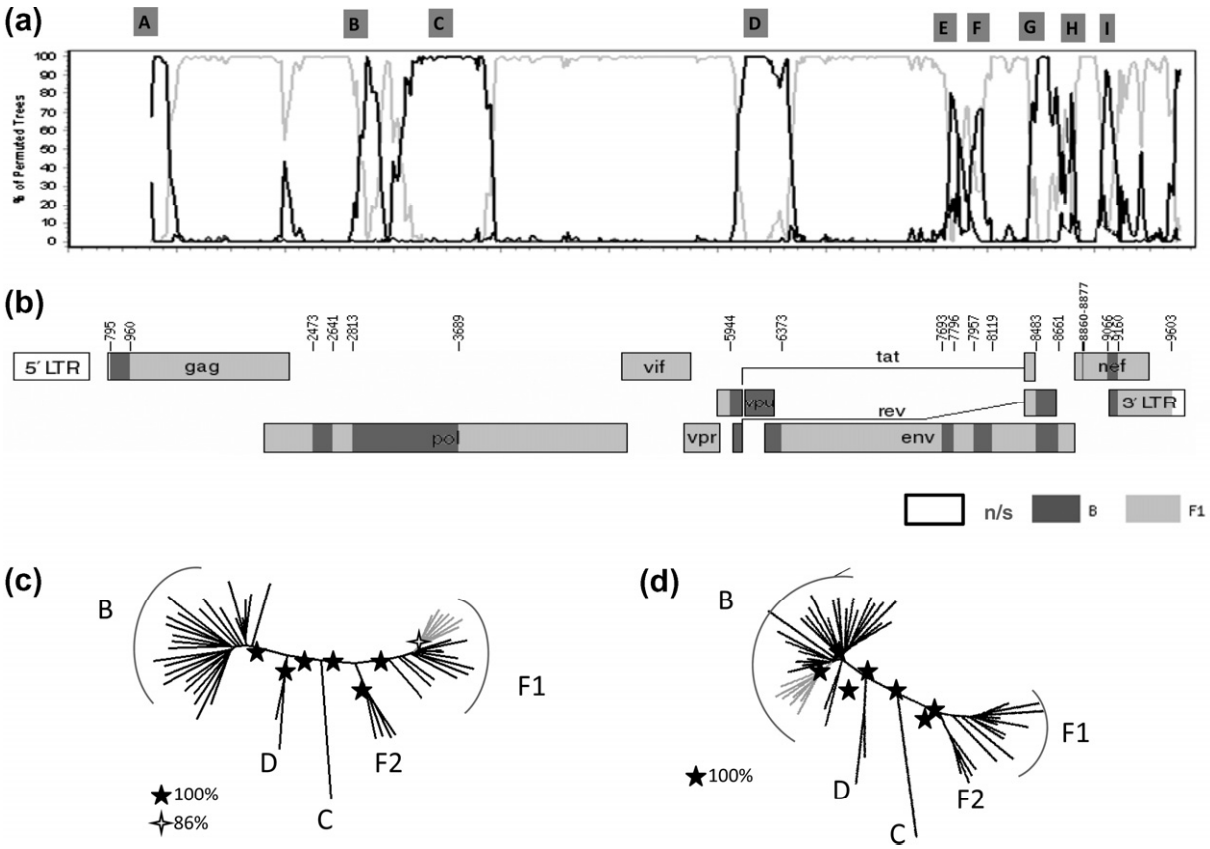

Fig. 2. Characterization of the recombinant structure of CRF17_BF. (a) Bootscanning plot of CRF17_BF ARMA038 strain, showing subtype B (black) and F1 (light gray) similarity along the full-length HIV-1 genome. In bootscan analyses performed with Simplot, bootstrap values were determined in N–J trees constructed using the Kimura two-parameter model, based on 100 re-samplings, supporting branching with the consensus sequences within a 200 bp window moving in steps of 20 bases. (b) Proposed recombination pattern of CRF17_BF based on the bootscanning plot, subtype signature patterns, and phylogenetic reconstruction of subtype-delimited segments. n/s: not sequenced. (c) Maximum likelihood phylogenetic tree of concatenated subtype F segments defined for CRF17_BF. (d) Maximum likelihood phylogenetic tree of concatenated subtype B segments defined for CRF17_BF. Star symbols represent percentage bootstrap values supporting the clades. Gray branches represent CRF17_BF viruses.

3.3. Comparison between CRF17_BF and CRF12_BF recombination profiles

A comparison between the recombination patterns of CRF12_BF -the prototype BF recombinant strain- and the newly characterized CRF17_BF strains, shows that CRF12_BF shares eight of the eighteen breakpoints defined for CRF17_BF (Fig. 3a), indicating coincident but independent BF recombination events on hot-spot regions of the HIV-1 genome. While subtype B segments identified with letters A, B, and G share the exact same recombination breakpoints in both CRF17_BF and CRF12_BF, segments C/c and D/d share only the 3′ or the 5′ recombination breakpoints, respectively among both CRFs. Segments E, F, H and I, were exclusively found in CRF17_BF, suggesting independent evolution of the parental strains or additional recombination events on a CRF12_BF background.

The phylogenetic relationship of each putative subtype B segment (A–I) delimited for CRF17_BF strains was analyzed and compared to CRF12_BF and subtype B and F1 reference strains through NJ phylogenetic trees (Fig. 3b). CRF17_BF strains clustered with subtype B references with high bootstrap support for segments A, B, C, D, and G, confirming the subtype B assignment by bootscans, signature nucleotide analysis, and RIP. The CRF12_BF strains clustered away from CRF17_BF for segments C and D, and were intermingled with CRF17_BF strains for segments A, B and G. Phylogenetic analysis of regions defined as segments E, F, and I, showed that CRF17_BF clustered within the subtype B group of reference strains, while CRF12_BF clustered with subtype F1 strains despite the bootstrap support was not high enough to define well separated subtype B and F1 clades.

4. Discussion

The Los Alamos HIV-1 Sequence Database (HIV Sequence Database, http://www.hiv.lanl.gov) provides an up-to-date collection of HIV-1 CRFs based on individual reports by the authors, usually from a set of at least two full-length and one partial sequence. CRF17_BF was reported as a single sequence (URF) in 2001, and only in 2008 six new sequences with a similar recombination pattern became available. For the first time, we confirm CRF17_BF as a genuine CRF based on the analysis of seven full-length sequences from four different countries (Argentina, Bolivia, Paraguay and Peru). Our analyses indicate that the recombination pattern of CRF17_BF is more complex than initially described (Carr et al., 2001), but continues to share several recombination breakpoints with the CRF12_BF, suggesting that both CRFs have evolved from a common ancestral BF recombinant.

To date, 49 different HIV-1 CRFs and one HIV-2 CRF have been described worldwide. Of them, the most frequent combination of subtypes results in 11 different BF recombinants representing 22% of all the CRFs (HIV Sequence Database, http://www.hiv.lanl.gov). Despite their over-abundance in comparison to other combinations of subtypes, BF recombinants are almost exclusively found in South-American countries (HIV Sequence Database, http://www.hiv.lanl.gov) or in European countries with a high social and cultural exchange with Latin-America like Spain and Italy (Fernández-García et al., 2010). CRF12_BF is considered the prototype of the BF recombinants, not only because it was the first to be described in 2001, but also because of its high abundance (Carr et al., 2001). CRF17_BF, on the contrary, seems to circulate at very low prevalence in Latin America. Up to the present, CRF17_BF was

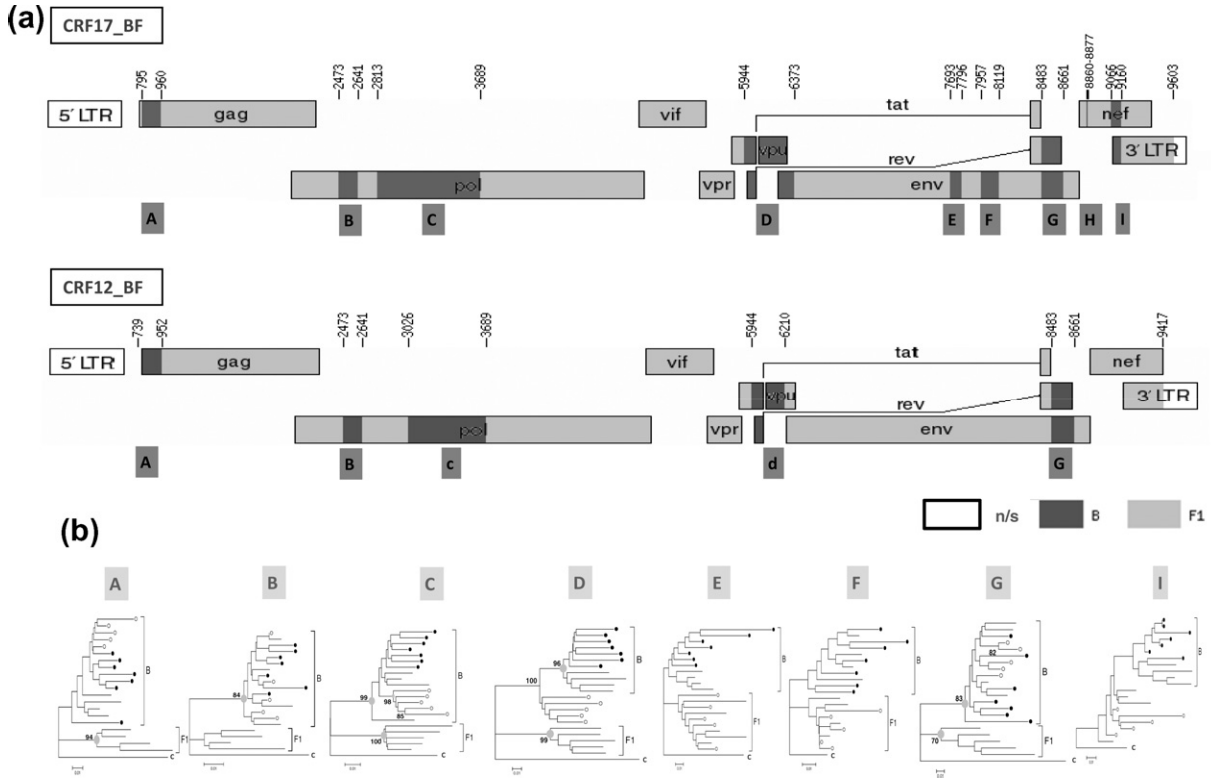

Fig. 3. Comparison of BF mosaic structures of CRF17_BF and CRF12_BF. (a) Comparison between the mosaic structure of CRF17_BF and CRF12_BF. Uppercase and lowercase letters indicate similar but not identical subtype B segments. n/s: not sequenced. (b) Neighbor-joining phylogenetic trees of individual CRF17_BF subtype B fragments A, B, C, D, E, F, G, and I (phylogenetic reconstruction of segment H is not depicted due to unresolved topology for subtype B and F1 references). Clades with bootstrap support above 70% are shown on the nodes. CRF17_BF sequences are represented with a full circle, and CRF12_BF sequences with an open circle. Subtype B segments are identified with a letter from A to I.

found in less than 2% of HIV-1 BF recombinants in Argentina (Aulicino et al., 2011), 0.4% of the HIV-1 strains circulating in South-America (Laguna-Torres et al., 2005), and 2% of the HIV-1 strains in Paraguay (Aguayo et al., 2008).

The recombination pattern of the CRF17_BF was originally described to be very similar to that of CRF12_BF. Only two differences in the length of the subtype B segment in *vpu* and one additional subtype B insertion in 5′ *env* gene of CRF17_BF were noted to distinguish between both CRFs (Carr et al., 2001). However, here we show that CRF12_BF and CRF17_BF have a similar mosaic structure but can be differentiated in their recombination pattern in several genomic regions including *vpu*, *pol*, *env* and *nef*. In fact, the differences in *pol* recombinant structures were recently used to characterize the BF recombinant epidemic in a group of children from Argentina, allowing for the first time to determine the frequency of CRF17_BF-like *pol* genotypes in this country (Aulicino et al., 2011).

It has been shown that the choice of window size affects the sensitivity of the detection of recombinants. On the one hand, using a small window size may introduce artifacts (small regions that appear to be of another subtype, but are not). On the other hand, using an overly-large window size may mask the presence of legitimate regions of recombination. Different sliding window and step sizes, as well as inappropriate subtype references often result in discordant recombination maps and affect the tracking of the epidemic of HIV-1 recombinants. Here we used a sliding window of 200 bp, and a step size of 20, and corroborated the results obtained by bootscanning with other methods to build the proposed CRF17_BF recombinant structure. In bootscans, we set window and step sizes in between the ones used for CRF44_BF (window: 160 bp/step: 10) (Delgado et al., 2010) and CRF38_BF (window: 300 bp/step: 30) (Ruchansky et al., 2009), two CRFs with a reported ancestral relationship to CRF12_BF and CRF17_BF (Sierra et al., 2005). Therefore, the use of a different set of parameters may account for differences between the previous CRF17_BF structure and the new one here presented. In fact, the sensitivity of detection of small subtype B segments E and H in CRF17_BF is lost for windows above 250 bp (data not shown).

Previous studies have shown that recombination patterns are often subject to re-analysis and change, demonstrating that our ability to pinpoint exactly the recombination breakpoints in any virus or group of viruses is not complete (Magiorkinis et al., 2005; Zhang et al., 2008, 2010). In particular, the detection of small recombinant segments represents a problem unsolved by the methods that require a minimum window size to build reliable trees and only allow one window size at a time, since it has been shown that the strength of the phylogenetic signal varies over the length of the HIV-1 genome (Hraber et al., 2008). In our study, we consider that a high density of informative sites and subtype signature nucleotides strong enough to cause a "blip" in bootscan analyses is more probably a result of recombination than a consequence of convergent evolution. In spite of the limitations imposed by the methods and sequences available, a complete and accurate characterization of recombination breakpoints in CRFs is required for a proper surveillance of HIV-1 genotypes in regions like South-America where the diversity of BF recombinants is increasing every day.

Acknowledgments

The authors wish to thank the 16th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology for fostering this collaboration.

References

Aguayo, N., Laguna-Torres, V.A., Villafane, M., Barboza, A., Sosa, L., Chauca, G., Carrion, G., Coenca, B., Perez, J., Galeano, A., Bautista, C.T., Sanchez, J.L., Carr, J.K., Kochel, T., 2008. Epidemiological and molecular characteristics of HIV-1 infection among female commercial sex workers, men who have sex with men and people living with AIDS in Paraguay. Rev. Soc. Bas. Med. Trop. 41, 225–231.

Aulicino, P.C., Bello, G., Guimaraes, M.L., Ruchansky, D., Rocco, C., Mangano, A., Morgado, M.G., Sen, L., 2011. Longitudinal analysis of HIV-1 BF1 recombinant strains in vertically infected children from Argentina reveals a decrease in CRF12_BF pol gene mosaic patterns and high diversity of BF unique recombinant forms. Infect. Genet. Evol. 11, 349–357.

Carr, J.K., Avila, M., Gomez Carrillo, M., Salomon, H., Hierholzer, J., Watanaveeradej, V., Pando, M.A., Negrete, M., Russell, K.L., Sanchez, J., Birx, D.L., Andrade, R., Vinoles, J., McCutchan, F.E., 2001. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. AIDS 15, F41–F47.

Delgado, E., Ríos, M., Fernández, J., Pérez-Alvarez, L., Nájera, R., Thomson, M.M., 2010. Identification of a new HIV Type 1 BF intersubtype circulating recombinant form (CRF44_BF) in Chile. AIDS Res. Hum. Retroviruses 26, 821–826.

Fernández-García, A., Pérez-Alvarez, L., Cuevas, M.T., Delgado, E., Muñoz-Nieto, M., Cilla, G., Iribarren, J.A., Pinilla, M., Ocampo, A., Miralles, C., Pérez-Castro, S., González-Galeano, M., de Castro, R.O., Trigo, M., García, V., Sánchez, A.M., Thomson, M.M., 2010. Identification of a new HIV type 1 circulating BF intersubtype recombinant form (CRF47_BF) in Spain. AIDS Res. Hum. Retroviruses 26, 827–832.

Hraber, P., Kuiken, C., Waugh, M., Geer, S., Bruno, W.J., Leitner, T., 2008. Classification of hepatitis C virus and human immunodeficiency virus-1 sequences with the branching índex. J. Gen. Virol. 89, 2098–2107.

Laguna-Torres, V.A., Olson, J., Sánchez, J.L., Montano, S., Chauca, G., Carrión, G., Romero, A., Ríos, J., Gamero, M.E., Sovero, M., Pérez-Bao, J., Carr, J.Grupo de Trabajo de Genotipificación de VIH en Sudamérica, 2005. Distribución de los subtipos del VIH-1 en nueve países de América del Sur, 1995–2002. Rev. Perú. Med. Exp. Salud Pública 22, 12–17.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.

Magiorkinis, G., Paraskevis, D., Magiorkinis, E., Vandamme, A.-M., Hatzakis, A., 2005. Re-analysis of 34 full-length HIV-1 intersubtype recombinant sequences. Infect. Genet. Evol. 5, 225–229.

Ruchansky, D., Casado, C., Russi, J.C., Arbiza, J.R., Lopez-Galindez, C., 2009. Identification of a new HIV Type 1 circulating recombinant form (CRF38_BF1) in Uruguay. AIDS Res. Hum. Retroviruses 25, 351–356.

Sierra, M., Thomson, M.M., Rios, M., Casado, G., Ojeda-de Castro, R., Delgado, E., Echevarría, G., Muñoz, M., Colomina, J., Carmona, R., Vega, Y., Vázquez –de Parga, E., Medrano, L., Pérez-Álvarez, L., Contreras, G., Nájera, R., 2005. The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. Infect. Genet. Evol. 5, 209–217.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Zhang, C., Ding, N., Wei, J.F., 2008. Different sliding window sizes and inappropriate subtype references result in discordant mosaic maps and breakpoint locations of HIV-1 CRFs. Infect. Genet. Evol. 8, 693–697.

Zhang, C., Zhang, M., Foley, B., Schultz, A.K., Macke, J.P., Bulla, I., Stanke, M., Morgenstern, B., Korber, B., Leitner, T., 2010. The role of recombination in the emergence of a complex and dynamic HIV epidemic. Retrovirology 7, 25.