Evolutionary Dynamics of HIV-1 BF and CB Recombinants and Its Parental Counterparts in South America

Élcio Leal[1], Leonardo O. Martins[2], L. Mario Janini[1] and Ricardo S. Diaz[1]

[1]Federal University of São Paulo, São Paulo, Brazil. [2]Graduate School of Agriculture and Life Sciences, University of Tokyo.

Abstract: Analysis of the near full-length genomes revealed that the subtype F appeared in Brazilian HIV-1 epidemics at least 10 years after the subtype B. Notably, the BF recombinant emerged almost simultaneously with the introduction of subtype F in Brazil. Analysis of reverse transcriptase fragments indicated that the C subtype originated in the early 1990s, and the CB recombinant emerged 2 years after the appearance of subtype C. The high growth rate of BF recombinant possibly obscured the prevalence of the pure subtype F. In contrast, subtype C, although appearing 20 years after subtype B, was responsible for a well-defined epidemic. Nevertheless, the CB recombinant equally emerged rapidly after the introduction of the second parental (subtype C). Our results suggest that the outcome related to the recombinant profile are probably influenced by the capacity of the newly arriving subtype to establish a critical number of infections before it recombines with the previous circulating subtype.

Keywords: HIV-1, South America, recombination, coalescent

Introduction

Circulating recombinant forms of HIV-1 (CRF) are particular intersubtypes recombinant viruses that spread throughout a population of epidemiologically unrelated individuals. Despite the fact that subtype B remains the most prevalent clade in South America, BF recombinant viruses are co-circulating in South American countries (Sierra et al. 2005; Santos et al. 2006; Bello et al. 2007). Previous studies have suggested that BF viruses circulating in Argentina (CRF_12) emerged from a single BF ancestral sequence originating in Brazil.(Sierra, Thomson et al. 2005) Although this might be true, the evolutionary dynamics and the dates of the original BF recombination events are less than clear. In HIV-1 epidemics occurring in the southern states of Brazil, nearly half of the cases involve infection with subtype C. (Soares et al. 2003) The complexity of subtype distribution and mosaic structures might have derived from an intricate combination of factors, including highly interconnected transmission networks exploited by distinct variants and the intrinsic recombination rate of HIV. A phylogenetic-based analysis using Bayesian methods was employed to explore the evolutionary history of the subtypes B, F and C, as well of the intersubtype recombinants BF and CB, the last having been identified only recently. (Santos, Sousa et al. 2006).

Our primary focus was the Brazilian epidemic, which seems to be the main source of the wider South American epidemic. Two distinct Bayesian coalescent methods were used to analyze near full-length sequences of pure B, F and C subtypes, as well as of the BF and CB recombinants. We also analyzed small fragments of the reverse transcriptase genes of the pure subtype C and of the recently isolated recombinant CB. We found that recombinant viruses (i.e. BF and CB) emerged just after the introduction of the second parental, either F or C subtype, into the Brazilian HIV-1 epidemics. Therefore, the number of infections caused by newly introduced subtype before it recombines with the previous circulating subtype might be decisive to determine the mosaic pattern of recombinant viruses.

Methods

Data sets compilation

We collected sequences from unlinked individuals sampled over distinct years (heterochronous) as follows: nine near full-length subtype B sequences of HIV-1 isolated and 45 near full-length BF virus

Correspondence: Élcio Leal, Federal University of São Paulo, Rua Pedro de Toledo, 781—16º andar—Vila Clementino—São Paulo—SP—Brazil, CEP 04039 032. Tel/Fax: (55 11) 5081 5394/5571 2130/5579 8226; Email: e.leal@unifesp.br

sequences of HIV-1, all isolated in South America. Five near full-length subtype F virus sequences, isolated in Brazil, were also analyzed. In addition, we analyzed 488-bp *env* fragments, obtained from 19 samples of subtype B circulating in South America. We also collected sequences sampled at same year (isochronous) as follows: 948-bp of *pol* gene sequences of subtype B from 126 Brazilian samples and 148 Argentine samples. We also analyzed a fragment of 579-bp RT gene of 43 samples of subtype C and of 32 samples of the newly identified CB recombinant viruses, all of which were obtained from cases occurring in Southern Brazil. (Soares, De Oliveira et al. 2003; Santos, Sousa et al. 2006).

All sequences included in this study were obtained from the GenBank. Detailed descriptions of these sequences are available in Supplementary information (Table S1).

Sequence alignments

The sequences were first aligned using the ClustalX program (Thompson et al. 1997). All sites with deletions and insertions were then excluded in order to preserve the reading frames of the genes. After this editing process, the sequences were manually aligned using the SE-AL program, version 2.0 (Department of Zoology, Oxford University; http://evolve.zoo.ox.ac.uk/software). Before the analysis the following drug-resistance codons were removed, according to the Stanford University HIV Drug Resistance Database (http://hivdb.stanford.edu): R8, L10, L23, V32, G48, I50, V82, I84 and L90 from protease; and M41, E44, K65, D67, T69, K70, L74, V75, L92, A98, L100, K101, K103, V106, V108, Y115, V118, E138, T139, G141, Q151, Q161, V179, Y181, M184, Y187, V188, G190, H208, L210, T215 and K219 from reverse transcriptase (RT).

Phylogenetic inferences

Maximum likelihood trees were done under the GTR model (Rodriguez et al. 1990) plus a Gamma distribution (GTR+Γ). This model was used to obtain an initial tree through the neighbor-joining method and then it was swapped by nearest-neighbor interchange method (NNI). Phylogenies used on coalescent analysis were made assuming maximum likelihood criteria and molecular clock. The inferences were made using the program Hyphy. (Pond et al. 2005).

Recombination detection

Since recombination can increase the overall length of terminal branches in phylogenies and this can affect the estimation of evolutionary parameters, such as substitution rates and time to the most common recent ancestor (Schierup et al. 2000). We then decided to detect and exclude recombinant sequences from the alignments used in the analyses. Likewise, in the analyses where recombinant viruses were evaluated we excluded DNA fragments with breakpoints. Description of breakpoints pattern of South American BF viruses can bee seen in the figure 1S (supplementary data). Detection of recombinant sequences and identification of breakpoints were done using DualBrothers software (Minin et al. 2005).

Coalescent analysis

A Bayesian Markov chain Monte Carlo (BMCMC) method was used. (Drummond et al. 2007). The BMCMC method uses the phylogenetic uncertainty of heterochronous sequences to estimate ancestral genealogies, demographic variables and evolutionary parameters such as nucleotide substitution rates per year, growth rate and divergence dates. We assumed relaxed molecular clock and tested three demographic models: (i) constant population size, (ii) exponential growth and (iii) logistic growth, characterized by a period of exponential growth followed by a decline. The MCMC chain length was 60–100 10^4 (with 10% of burn-in). The analyses were made using the BEAST package, and the convergence of parameters was evaluated using TRACER software, version 1.4 (Department of Zoology, Oxford University (http://evolve.zoo.ox.ac.uk/software).

Additionally, a strict molecular clock technique, based on the reversible-jump MCMC method (rjMCMC) implemented in the "ape" library of the R-package (http://www.r-project.org) (Opgen-Rhein et al. 2005), was used to analyze the demographic history of HIV from fixed binary trees constructed with isochronous sequences. The coalescent analyses were performed according to the following criteria: First we applied the MCMC method to analyze nine near full-length subtype B sequences isolated in South America and five full-length subtype F sequences isolated in Brazil. This technique was also applied to a 488-bp *env* fragment from 19 samples of HIV-1 of subtype B from Brazil.

Finally, a 1258-bp *env* fragment, free of breakpoints and with a genetic composition related to the subtype F (herein denoted clade F), obtained from 26 sequences of BF virus was also analyzed. Additionally, the rjMCMC method was applied to isochronously-sampled sequences. Prior to the analyses, sequences were used to construct phylogenetic trees assuming a strict molecular clock. Those trees were then used to determine the demographic history through coalescent-based measurement of the internode intervals, resulting in a skyline plot, in which the effective population size is depicted along the time axis. We analyzed 948-bp *pol* sequences of subtype B: 126 Brazilian samples and 148 Argentine samples. The same approach was also used to analyze a 1210-bp *env* fragment (herein denominated clade B), free of breakpoints and with a genetic composition related to the subtype B (monophyletic) obtained from 17 sequences of BF virus was also analyzed. Because these sequences were isolated in 1999 they were considered isochronous. Finally we analyzed 579-bp RT fragments of 43 subtype C samples and of 28 samples of the newly identified CB recombinant viruses, all of which were obtained from cases occurring in Southern Brazil. In order to reduce biases due to convergence, drug-resistance codons were removed from *pol* and RT sequences.

Results

Demographic history of subtype B in Brazil and in Argentina

Since the most probable origin for the CRF_12 was due to a unique transmission event from Brazil. We then evaluated the demographic history of subtype B strains in Brazil and Argentina. To perform this analysis, trees were constructed using the *pol* fragments of subtype B sequences from 126 Brazilian samples and 148 Argentine samples. Initially, a phylogenetic analysis was constructed in which we included samples from Brazil and Argentina plus reference isolates of HIV-1. This tree was used to detect recombinant sequences. All sequences presenting phylogenetic evidence of recombination were excluded, as were drug-resistance codons. Maximum likelihood trees were then constructed assuming a strict molecular clock and they were next used in the rjMCMC analysis. For both the Brazilian and the Argentine subtype B *pol* trees, a substitution rate (substitutions per site per year: μ) of 0.0015 (Salemi et al. 2005) was assumed. The skyline plot indicated that the epidemic growth of subtype B in Brazil increased exponentially after 1975 (Fig. 1, panel A) and in Argentina after 1970 (Fig. 1, panel B). This result might indicate that HIV-1 subtype B epidemic started almost in the same time in Brazil and Argentina and spread similarly in these countries.

Phylogenetic analysis of near full-length genomes

In order to determine the relationship of strains circulation in South America we constructed a maximum likelihood tree using near full-length sequences. The tree (not shown) has a monophyletic group formed by CRF_12 with few recombinants from Argentina dispersed in the tree. On the other hand, recombinants from Brazil were almost all scattered in the near full-length genome tree. The diffuse pattern of some recombinant sequences is expected since they present distinct mosaic patterns and might have originated from distinct ancestral sequences. By this reason we then decided to study a genomic fragment free of breakpoints that could reflect the ancestral relationship of recombinant viruses. We identified a 1258-bp *env* fragment absent of breakpoints for most recombinants from Argentina and for nine recombinants from Brazil (herein designated as clade F and clade B). This breakpoints-free *env* fragment was used to study the demographic history of BF strains from Argentina (clade F) and of BF strains from Brazil (clade B).

Demographic history of BF recombinants

Initially, the rjMCMC method was also used to analyze the *env* fragment from sequences of the BF recombinant of the Clade F. This fragment is also free of breakpoints, and its nucleotide composition is based upon the subtype F parental genome. In this case, sequences of the *env* fragment were assumed to be isochronous and a substitution rate of 0.009 (we estimated this rate using the relaxed molecular clock Bayesian MCMC method as is explained below) was used. The result shows that recombinants of clade F start to growth exponentially after 1986 (Fig. 2, panel A). Similarly, The rjMCMC method was

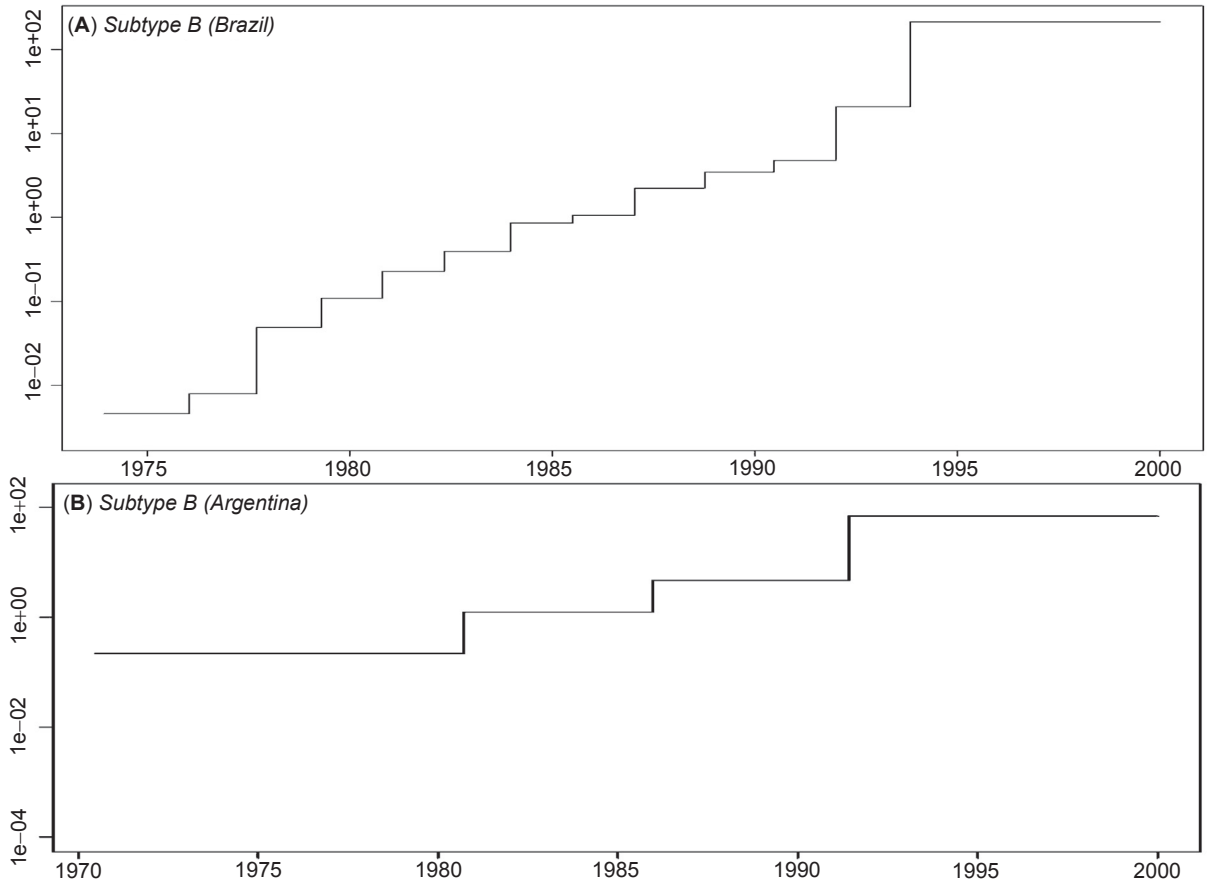

Figure 1. Skyline plot of HIV-1 subtype B from Brazil and Argentina. Panel A: tree constructed with 948-bp *pol* fragments of subtype B obtained from Brazil; Panel B: tree constructed with 948-bp *pol* fragments of subtype B obtained from Argentina. The following drug-resistance codons were removed, according to the Stanford University HIV Drug Resistance Database (http://hivdb.stanford.edu): R8, L10, L23, V32, G48, I50, V82, I84 and L90 from protease; and M41, E44, K65, D67, T69, K70, L74, V75, L92, A98, L100, K101, K103, V106, V108, Y115, V118, E138, T139, G141, Q151, Q161, V179, Y181, M184, Y187, V188, G190, H208, L210, T215 and K219 from reverse transcriptase (RT).

also used to analyze the 1210-bp *env* fragment (after further additional trimming) from sequences of the BF recombinant of the Clade B. This fragment is free of breakpoints, and its nucleotide composition is based upon the subtype B parental genome and a substitution rate of 0.009 was assumed. The result shows that recombinants of clade B started to increase exponentially before 1985 (Fig. 2, panel B).

Next, we used the relaxed molecular clock and the dates of sequences to estimate the demographic parameters of HIV-1 strains. The results of the heterochronous Bayesian MCMC analysis are summarized in Table 1. The demographic parameter estimates based on the near full-length genome showed that the age of the most recent ancestral (MRCA) of subtype B in South America originated in 1971 (1950–1982). On the other hand, the estimated age of MRCA of subtype F was 1981 (1968–1988) (Table 1). Other parameters, such as substitutions per site, growth rates and the number of infections, were lower for subtype F than for subtype B. This might reflect the minimal impact of subtype F on the HIV-1 epidemic in Brazil. In addition, the results obtained with the *env* fragment reveal that the MRCA of subtype B is 1970 (1940–1988) similar to the estimate made using the near full-length genome. Conversely, the *env* fragment estimates indicated a more recent origin for MRCA of BF viruses of clade F, dating from 1984 (1976–1988). Although these sequences were obtained from recombinant viruses, this fragment is free of breakpoints, and its nucleotide composition is based upon the subtype F parental genome. Notably, the estimated substitutions per site, growth rate and number of infections for BF viruses of clade F were higher than those for subtype F and similar to those for subtype B. Similarly, the

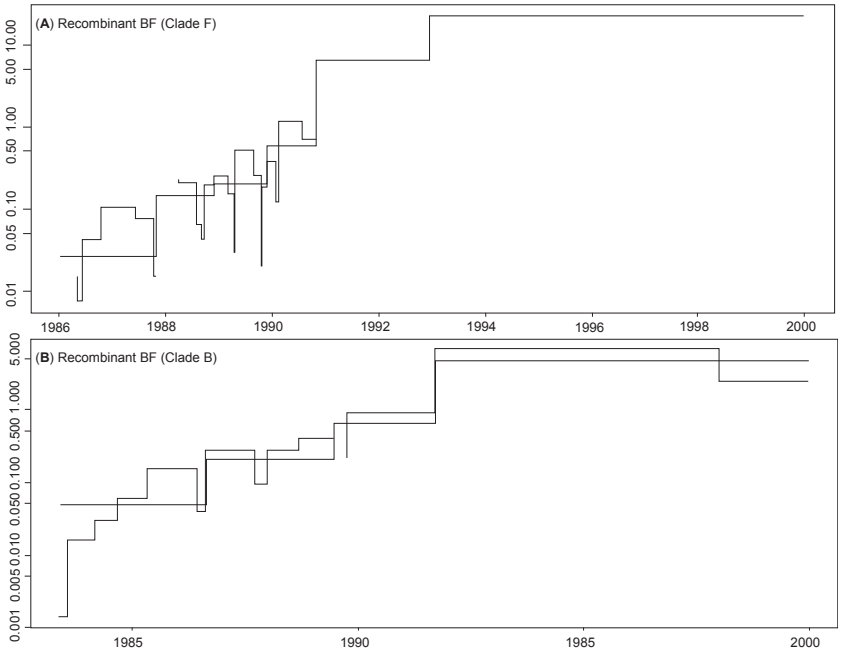

Figure 2. Skyline plot of BF recombinants of HIV-1 from South America. The Bayesian skyline plot depicting the coalescent intervals of each tree constructed under the molecular clock assumption. The *x*-axis is in units of time (years). The *y*-axis represents the effective population size. The grey line represents the classical skyline plot, and the dark line represents the generalized skyline plot. Panel A: tree constructed with the 1258-bp *env* fragment of clade F, free of breakpoints, obtained from BF recombinants isolated in South America; Panel B: tree constructed with the1210-bp *env* fragment of clade B, free of breakpoints, obtained from BF recombinants isolated in South America.

estimated MRCA of BF viruses of clade B, was 1989 (1984–1991). However other parameters such as substitutions per site, growth rate and number of infections were extremely high compared to the estimated for others viruses (Table 1).

Number of secondary infections

The growth rate (r) permits the estimation of the basic reproductive number of an infection (R_0), which reflects the number of new secondary infections caused by a primary infection pathogen. The number of new infections caused by HIV-1 can be estimated using the equation:

$$R_0 = rD + 1,$$

where D is the average duration of infectiousness. It is quite reasonable to assume a putative range of 5 to 10 years (Salemi, de Oliveira et al. 2005) for D for HIV-1, hence it provided an estimated mean R_0 of 3.107 (2.405 to 3.81) new secondary infections with subtype B using the growth rate estimated for the near full-length genome and mean R_0 of 2.9425 (2.295 to 3.59) using the growth rate estimated for the *env* gene fragment. The estimated mean R_0 was 2.395 (1.93 to 2.86) for subtype F (using the growth rate estimated for the near full-length genome). On the other hand, for BF recombinant the estimated mean R_0 was 3.6775 (2.785 to 4.57) for the clade F viruses and 6.025 (4.35 to 7.7) for clade B viruses (Table 2). These findings might reflect either the elevated transmission dynamics of or a selective advantage of BF viruses over others "pure" strains in the South American HIV-1 epidemic. Although we have used a limited number of sequences and may incur in biased estimates for all parameters, our results were in agreement with results observed in a cluster of BF recombinants isolated in children from Argentina (Aulicino et al. 2007).

Demographic parameters of CB recombinants

The rjMCMC method was applied to study the RT sequences of the subtype C and CB recombinant viruses. Two near full-length genomes of pure subtype C and two full-length genomes of the CB recombinant (Lole et al. 1999; Santos, Sousa et al. 2006) were initially used as references in a phylogenetic analysis to identify the RT sequences of subtype C and of the CB recombinant (Fig. 3). The tree identified two

Table 1. Evolutionary parameters of HIV-1 in South America.

Data set (n = Samples, l = Base pairs) [Date range]	Demographic model*	Mean substitutions/site year (μ)	Origin of the tree (year)	Mean growth rate/year	Number of infections (*Ne*)
		Near full-length genome			
B-SAmer (n = 9, l = 9853) [1989–2002]	Logistic	0.0054 (0.001–0.009)	1971 (1950–1982)	0.281 (0.106-0.43)	7816 (1562–13000)
F-Brazil (n = 5, l = 9853) [1989–2001]	Logistic	0.004 (0.001–0.006)	1981 (1968–1988)	0.186 (0.012–0.47)	2663 (230–10000)
		Gene fragment			
B-Brazil (n = 19, l = 488) [1989–2000]	Exponential	0.009 (0.002–0.016)	1970 (1940–1988)	0.259 (0.01–0.5)	5743 (3759–7000)
BF Clade-F (n = 18, l = 1258) [1994–1999]	Exponential	0.009 (0.005–0.015)	1984 (1974–1990)	0.357 (0.272–0.48)	6835 (1143–13000)
BF Clade-B (n = 9, l = 1210) [1999–2002]	Exponential	0.07 (0.022–0.085)	1989 (1984–1991)	0.670 (0.468–0.81)	6497 (1576–10000)
Cluster-C (n = 43, l = 579) [2000]	Exponential	0.0033 (0.003–0.0034)	1987 (1984–1989)	0.72 (0.49–0.971)	11500 (1085–53000)
Cluster-CB (n = 32, l = 579) [2000]	Exponential	0.0031 (0.003–0.0033)	1992 (1990–1992)	1.56 (0.91–2.35)	15020 (7896–93050)

*Best demographic model according to the Akaike information criterion, with the degrees of freedom given by the difference between each model in terms of the number of parameter.

Table 2. Basic reproductive number of secondary infections (R_0) in HIV-1 in South America.

Data set	Growth rate(*r*) (95% HPD interval)	R_0 with D = 5 years (interval)	R_0 with D = 10 years (interval)
B-SAmer (n = 9, l = 9853)	0.281 (0.106–0.43)	2.405 (1.53–3.15)	3.81 (2.06–5.3)
B-Brazil (n = 19, l = 488)	0.259 (0.01–0.5)	2.295 (1.05–3.5)	3.59 (1.1–6.0)
F-Brazil (n = 5, l = 9853)	0.186 (0.012–0.47)	1.93 (1.06–3.35)	2.84 (1.12–5.7)
BF Clade-F (n = 18, l = 1258)	0.357 (0.272–0.48)	2.785 (2.36–3.4)	4.57 (3.72–5.8)
BF Clade-B (n = 9, l = 1258)	0.670 (0.468–0.81)	4.35 (3.34–5.05)	7.7 (5.68–9.1)
Cluster-C (n = 43, l = 579) [2000]	0.72 (0.49–0.971)	4.6 (3.45–5.855)	8.2 (5.9–10.71)
Cluster-CB (n = 32, l = 579) [2000]	1.56 (0.91–2.35)	8.8 (5.55–12.75)	16.6 (10.1–24.5)

**D* average duration of infectiousness.

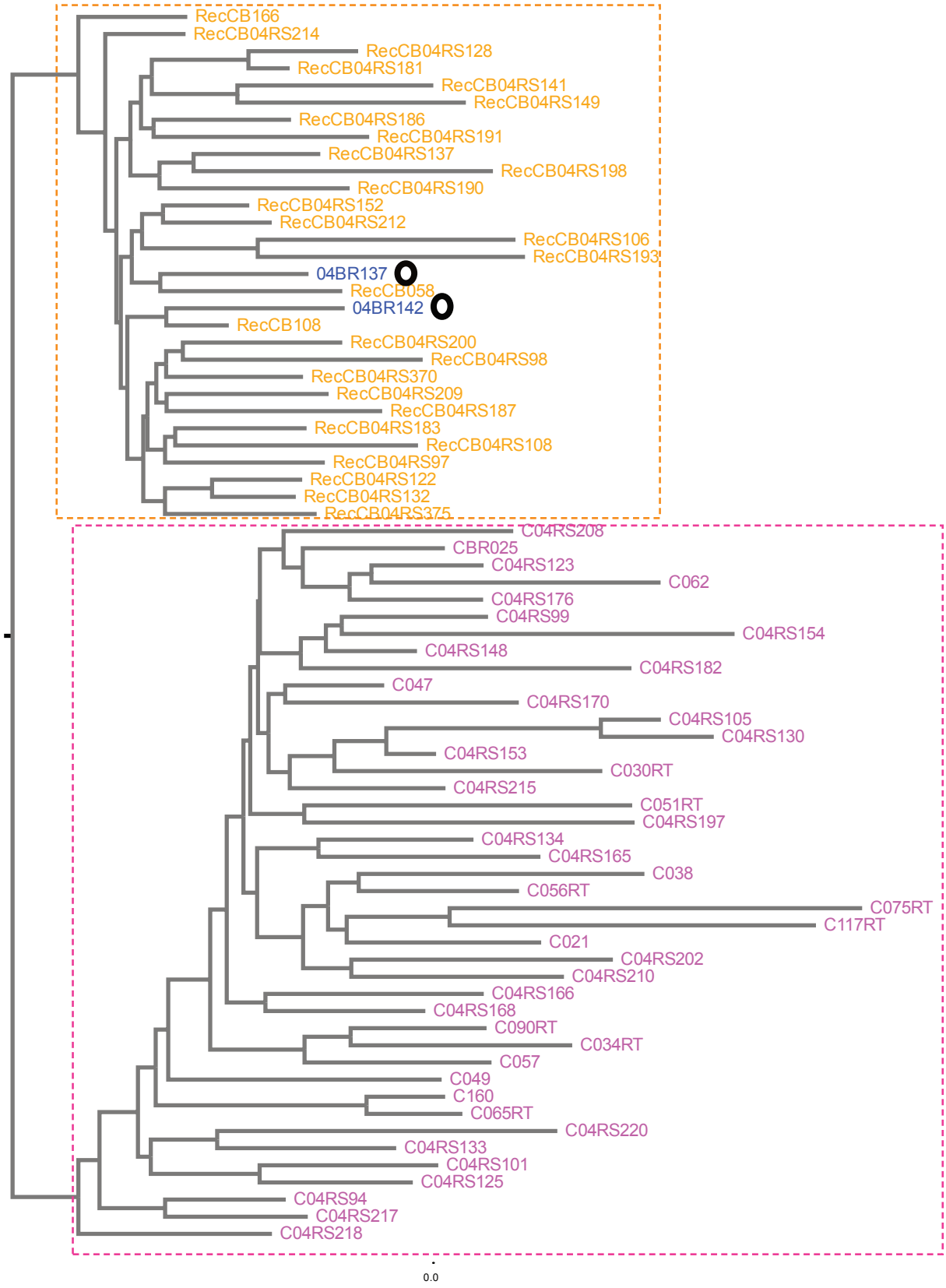

Figure 3. Maximum likelihood tree constructed using nearest-neighbor interchange method and 579-bp reverse transcriptase fragment of Brazilian isolates of HIV-1 of subtype C. Open circles depict two previous identified CB recombinants (04BR137 and 04BR colored in blue) which full genomes are available. They were used to identify monophyletic isolates (see text for details). The tree topology is composed by two cluster (delineated by dashed rectangles) that discriminate subtype C (sequences depicted in magenta) from the CB recombinant (sequences depicted in orange).

well-delineated clusters defining the subtype C and the monophyletic CB recombinant sequences. The CB sequences share an identical 259-bp B signature, suggesting that they derived from the same ancestor and were monophyletic in the tree. Separate strict molecular clock trees were constructed for the subtype C and CB recombinant sequences. These trees were employed in the rjMCMC analysis of the skyline plots, assuming a substitution rate of 0.00334 (Salemi, de Oliveira et al. 2005). The results show that subtype C dates back to 1990 (Fig. 4, panel A) whereas subtype CB is more recent, dating back to 1992 (Fig. 4, panel B). Additionally, we used the relaxed molecular clock assuming a substitution rate of 0.00334 (Salemi, de Oliveira et al. 2005) to estimate the demographic parameters of subtype C and CB recombinant viruses. The estimated MRCA for of subtype C is 1987 (1984–1989) and the estimated MRCA for CB recombinant virus is 1989 (1984–1991) (Table 1). Likewise, assuming a putative range of 5 to 10 years (Salemi, de Oliveira et al. 2005) for the duration of infectiousness it was possible to estimate the number of secondary infections (R_0) caused by subtype C and CB recombinants (Table 2). Therefore, the mean R_0 was 6.6 (4.6 to 8.2) for subtype C (using the growth rate of 0.72 (0.49–0.971) and mean R_0 was 12.7 (8.8 to 16.6) for CB recombinants (using the growth rate of 1.56 (0.91–2.35).

Discussion

The age of MRCA estimated for subtype B is in agreement with those of previous studies suggesting the 1970s as the most probable decade of origin of HIV-1 epidemics in Brazil. (Bello, Eyer-Silva et al. 2007) It also agrees with recent finding suggesting that HIV-1 spread as a single event from Haiti after the 70s. (Gilbert et al. 2007) The findings that subtype F and the BF recombinant emerged simultaneous in Brazil is also concordant with the identification of a BF virus in the mid-1980s (Carr et al. 2001) and of the first case of subtype F in the early 1990s (Potts et al. 1993). This finding, together with the lack of a pure subtype F epidemic in South American countries, strengthens the hypothesis that the now widely-disseminated BF circulating recombinant forms originated in Brazil (Sierra, Thomson et al. 2005). In addition, our findings suggest that subtype F was readily and completely assimilated into the previous (caused by subtype B) HIV-1 epidemics in Brazil. Perhaps the rapid assimilation of the pure subtype F has drastically reduced the chances of this subtype establishing a minimum number of infections and thus independently initiating an epidemic. This idea is also supported by the broad diversity of BF recombinant

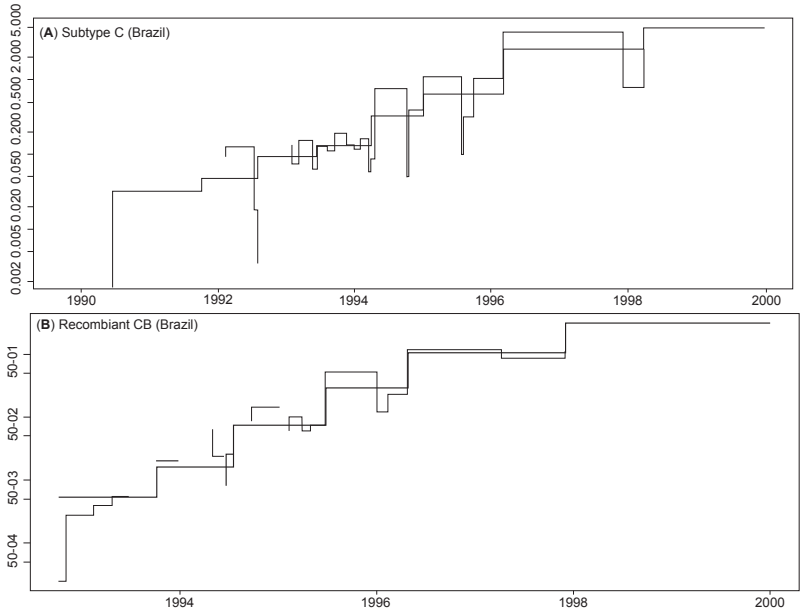

Figure 4. Skyline plot of HIV-1 subtype C and recombinant CB from Brazil. Panel A: skyline plot of the tree obtained with 579-bp reverse transcriptase fragments of 43 subtype C samples from Brazil; Panel B: skyline plot of the tree constructed with 28 samples of the CB recombinant viruses recently identified in Brazil.

forms (Gao et al. 1998; Thomson et al. 2002; Sa Filho et al. 2005; Sierra, Thomson et al. 2005), as well as the limited diversity of pure subtype F (Louwagie et al. 1994), observed in South America. On the other hand, results also show that, in Brazil, subtype C emerged 20 years after subtype B, coinciding with previous estimates (Salemi, de Oliveira et al. 2005). Furthermore, the CB recombinant virus emerged in Brazil just 2 years after subtype C, thus indicating that recombination between parental subtypes also occurred immediately after the first intermingling. However, in this case the subtype CB assimilated only a small fragment of the parental subtype B. It may indicate some epidemiological property of subtype C viruses since this subtype apparently is replacing the subtype B in Southern areas of Brazil. (Soares, De Oliveira et al. 2003; Salemi, de Oliveira et al. 2005).

The rapid and complete assimilation of subtype F resulted in the emergence of distinct BF viruses with varying proportions of B and F fragments along their genome. In addition, the complex spectrum of unique recombinant forms might also result from *de novo* recombination. Consequently, the complete assimilation of subtype F might have drastically reduced the chances for this subtype to establish an epidemic. Nevertheless, this hypothesis might not hold for other subtypes, since the CB recombinant also emerged simultaneously with subtype C in Brazil. Distinctly, subtype C is growing more rapidly than is subtype B and accounts for approximately half of the HIV-1 cases in Southern Brazil (Soares, De Oliveira et al. 2003; Salemi, de Oliveira et al. 2005). Notably, the recombinant form resulted from the intermingling of subtypes B and C presents an RT gene parental B fragment of less than 300 bp. Therefore, subtype C supposedly presents better fitness than does subtype B. The biological and epidemiological reasons for this distinction in the way subtypes F and C were assimilated into the HIV-1 epidemic in Brazil merits further study. Furthermore, the high growth rate and consequent high R_0 of recombinants might indicate that it presents greater fitness than do its parental counterparts. However, another intriguing feature of BF recombinant epidemics is the way they spread in Brazil and Argentina. Particularly, although CRF-12 may have its origin in Brazil they are not disseminated epidemically in Brazil. In addition, nearly 80% of all HIV-1 isolates from Argentina are BF recombinants of clade F (i.e. CRF-12 and others recombinants with similar mosaic pattern). (Sierra, Thomson et al. 2005; Aulicino, Holmes et al. 2007) Since BF recombinants from Argentina have extremely high growth rates (Aulicino, Holmes et al. 2007). Therefore it is quite reasonable to suggest that the elevated prevalence of BF recombinants among Argentinean isolates is due to the replicative fitness of these viruses. Nevertheless, most of BF viruses lately isolated in Brazil which includes two CRFs, (see Fig. 1S or ref. (Sa Filho, Sanabani et al. 2005) for a detailed description of these CRFs) have its genome composed primarily by subtype B background and our study might indicate that they also have high replicative rates. Therefore the reason for the successful spread of CRF-12 only in Argentina and not in Brazil cannot be explained merely by its supposed elevated replicative fitness. Consequently, it is tempting to speculate that the key for the successful dissemination of an HIV-1 clade (either pure subtypes or recombinants) relies in first establishing a critical number of infections in moderately isolated groups of highly interconnected individuals, such as intravenous drug users. Hence, this temporary isolation may provide the critical number of infected hosts needed for viruses to grow enough and establish a characterized epidemic that subsequently enable the distinct virus to spread to other individuals. Indeed this also seems be the case for subtype B since the early spread of almost all isolates this subtype can be traced back to only one Haitian isolate. (Gilbert, Rambaut et al. 2007).

Acknowledgements

E.L. is supported by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, Foundation for the Support of Research in the State of São Paulo; grants no. 04/10372-3, 07/52841-8).

Disclosure

The authors report no conflicts of interest.

References

Aulicino, P.C., Holmes, E.C. et al. 2007. Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina. *J. Virol.*, 81:427–9.

Bello, G., Eyer-Silva, W.A. et al. 2007. Demographic history of HIV-1 subtypes B. and F in Brazil. *Infect. Genet. Evol.*, 7:263–70.

Carr, J.K., Avila, M. et al. 2001. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS*, 15:F41–7.

Drummond, A.J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.*, 7:214.

Gao, F., Robertson, D.L. et al. 1998. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.*, 72:5680–98.

Gilbert, M.T., Rambaut, A. et al. 2007. The emergence of HIV/AIDS in the Americas and beyond. *Proc. Natl. Acad. Sci. U.S.A.*, 104:18566–70.

Lole, K.S., Bollinger, R.C. et al. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.*, 73:152–60.

Louwagie, J., Delwart, E.L. et al. 1994. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res. Hum. Retroviruses*, 10:561–7.

Minin, V.N., Dorman, K.S. et al. 2005. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics*, 21:3034–42.

Opgen-Rhein, R., Fahrmeir, L. et al. 2005. Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evol. Biol.*, 5:6.

Pond, S.L., Frost, S.D. et al. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21:676–9.

Potts, K.E., Kalish, M.L. et al. 1993. Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. Brazilian Collaborative AIDS Research Group. *AIDS*, 7:1191–7.

Rodriguez, F., Oliver, J.L. et al. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.*, 142:485–501.

Sa Filho, D.J., Sanabani, S. et al. 2005. Analysis of full-length human immunodeficiency virus type 1 genome reveals a variable spectrum of subtypes B. and f recombinants in Sao Paulo, Brazil. *AIDS Res. Hum. Retroviruses*, 21:145–51.

Salemi, M., de Oliveira, T. et al. 2005. Different epidemic potentials of the HIV-1B and C subtypes. *J. Mol. Evol.*, 60:598–605.

Santos, A.F., Sousa, T.M. et al. 2006. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS*, 20:2011–9.

Schierup, M.H. and Hein, J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics*, 156:879–91.

Sierra, M., Thomson, M.M. et al. 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.*, 5:209–17.

Soares, M.A., De Oliveira, T. et al. 2003. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. *AIDS*, 17:11–21.

Thompson, J.D., Gibson, T.J. et al. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic. Acids. Res.*, 25:4876–82.

Thomson, M.M., Delgado, E. et al. 2002. 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. *J. Gen. Virol.*, 83:107–19.

Evolutionary Dynamics of HIV-1 BF and CB Recombinants and Its Parental Counterparts in South America

Élcio Leal, Leonardo O. Martins, L. Mario Janini and Ricardo S. Diaz

Supplementary Material

Table S1. Description of sequences of HIV-1 from South America.

Data set (n = Samples, l = Base pairs) [subtype]	Genomic region	GenBank accession number (Year of isolation)	Geographic location
		Heterochronous	
B-SAmer. (n = 9, l = 9853) [B]	Near full length	AY173956 (1989), AY037268 (1998), AY037282 (1999), AY037270 (1999), AY037269 (2000), DQ358809 (2002), DQ358805 (2002), DQ358808 (2002)	Brazil, Argentina, Bolivia
F-Brazil (n = 5, l = 9367 [F]	Near full length	AY173957 (1989), AY173958 (1989), AF005494 (1993), DQ358801 (2001), DQ358802 (2001),	Brazil
B-Brazil. (n = 19, l = 390) [B]	envelope (V3 loop)	AY736820 (1989), AY877001 (1992), AF060955 (1995), AF060957 (1995), AY877010 (1995), AY877007 (1995), AY876999 (1996), AY877011 (1997), AY877012 (1997), AY877014 (1997), AY877013 (1997), AY877016 (1999), AY877004 (1999), AY877020 (1999), AY8770201 (1999) AY877001 (2000), AY877017 (2000), AY877019 (2000), AF034032 (2000)	Brazil
BF-SAmer (n = 43, l = 9853) [BF]	Near full length	clade F: AY455781 (1994), (AY455784 (1994), AY455785 (1994), AY455779 (1994), AF408631 (1997), AF408629 (1997), AY037266 (1998), AF408627 (1999), AF385935 (1999), AF385936 (1999), AF408630(1997), AY491031 (1999), AY037280 (1999), AF385934 (1999), AF408632 (1999), AY037283 (1999), AF332867 (1999), AY037272 (1999), AY536238 (1999), AY037271 (1999), AY037281 (1999), AY455782 (1999), AY536234 (1999), AY536233 (1999), AY536235 (1999), AY536237 (1999). clade B (1999): AF408628 (1999), AY771592 (1999), AY771593 (1999), AY771589 (1999), AY771590 (1999), AY771588 (1999), AY771591 (1999), DQ085876 (1999), DQ085871 (1999), AY455778 (1999), DQ085869 (1999), DQ085870 (1999), DQ085872 (1999), DQ085874 (1999), DQ085873 (2000), DQ085875 (2002), AY536236 (2000)	Brazil, Argentina, Chile, Venezuela

(*Continued*)

Table S1. (*Continued*)

Data set (n = Samples, l = Base pairs) [subtype]	Genomic region	GenBank accession number (Year of isolation)	Geographic location
		Isochronous	
B-Brazil (n = 126, l = 948) [B]	polimerase	AY771375, AY771376, AY771377, AY771378, AY771379, AY771380, AY771381, AY771382, AY771383, AY771384, AY771385, AY771386, AY771387, AY771388, AY771389, AY771390, AY771391, AY771392, AY771393, AY771394, AY771395, AY771396, AY771397, AY771398, AY771399, AY771400, AY771401, AY771402, AY771403, AY771404, AY771405, AY771406, AY771407, AY771408, AY771409, AY771410, AY771411, AY771412, AY771413, AY771414, AY771415, AY771416, AY771417, AY771418, AY771419, AY771420, AY771421, AY771422, AY771423, AY771424, AY771425, AY771426, AY771427, AY771428, AY771429, AY771430, AY771431, AY771432, AY771433, AY771434, AY771435, AY771436, AY771437, AY771438, AY771439, AY771440, AY771441, AY771442, AY771443, AY771444, AY771445, AY771446, AY771447, AY771448, AY771449, AY771450, AY771451, AY771452, AY771453, AY771454, AY771455, AY771456, AY771457, AY771458, AY771459, AY771460, AY771461, AY771462, AY771463, AY771464, AY771465, AY771466, AY771467, AY771468, AY771469, AY771470, AY771471, AY771472, AY771473, AY771474, AY771475, AY771476, AY771477, AY771478, AY771479, AY771480, AY771481, AY771482, AY771483, AY771484, AY771485, AY771486, AY771487, AY771488, AY771489, AY771490, AY771491, AY771492, AY771493, AY771494, AY771495, AY771496, AY771497, AY771498, AY771499, AY771500	Brazil
B-Argentina (n = 148, l = 948) [B]	polimerase	AF456535, AF456536, AF456537, AF456538, AF456539, AF456540, AF456541, AF456542, AF456543, AF456544, AF456545, AF456546, AF456547, AF456548, AF456549, AF456550, AF456551, AF456552, AF456553, AF456554, AF456555, AF456556, AF456557, AF456558, AF456559, AF456560, AF456561, AF456562, AF456563, AF456564,	Argentina

(*Continued*)

Table S1. (*Continued*)

Data set (n = Samples, l = Base pairs) [subtype]	Genomic region	GenBank accession number (Year of isolation)	Geographic location
		AF456565, AF456566, AF456567, AF456571, AF456572, AF456573, AF456574, AF456575, AF456576, AF456577, AF456578, AF456579, AF456580, AF456581, AF456582, AF456583, AF456584, AF456585, AF456586, AF456587, AF456588, AF456589, AF456590, AF456591, AF456592, AF456593, AF456594, AF456595, AF456596, AF456597, AF456598, AF456599, AF456600, AF456601, AF456602, AF456603, AF456604, AF456605, AF456606, AF456607, AF456608, AF456609, AF456610, AF456611, AF456612, AF456613, AF456614, AF456615, AF456616, AF456617, AF456618, AF456619, AF456620, AF456621, AF456622, AF456623 , AF456624 , AF456625 , AF456626 , AF456627 , AF456628 , AF456629 ,AF456630 , AF456631	
C-Brazil (n = 43, l = 578) [C]	Reverse transcriptase	DQ070626, DQ070627, DQ070628, DQ070629, DQ070630, DQ070631, DQ070632, DQ070636, DQ070637, DQ070638, DQ070639, DQ070640, DQ070641, DQ070642, DQ070643, DQ070644, DQ070645, DQ070646, DQ070647, DQ070652, DQ070653, DQ070654, DQ070688, DQ070689, DQ070690, DQ070691, DQ070692, DQ070693, DQ070694, DQ070695, DQ070696, DQ070697, DQ070698, DQ070699, DQ070702, DQ070703, DQ070704, DQ070705, DQ070706, DQ070707, DQ070708, DQ070709, DQ070714	Brazil
CB-Brazil (n = 28, l = 578) [CB]	Reverse transcriptase	AY727527, AY275747, DQ070670, DQ070678, DQ070668, DQ070656, DQ070681, DQ070658, DQ070667, DQ070663, DQ070673, DQ070676, DQ070661, DQ070680, DQ070662, DQ070671, DQ070672, DQ070666, DQ070673, DQ070674, DQ070669, AY727526, DQ191004, DQ191007, DQ070664, DQ070665, DQ070675, DQ070679	Brazil

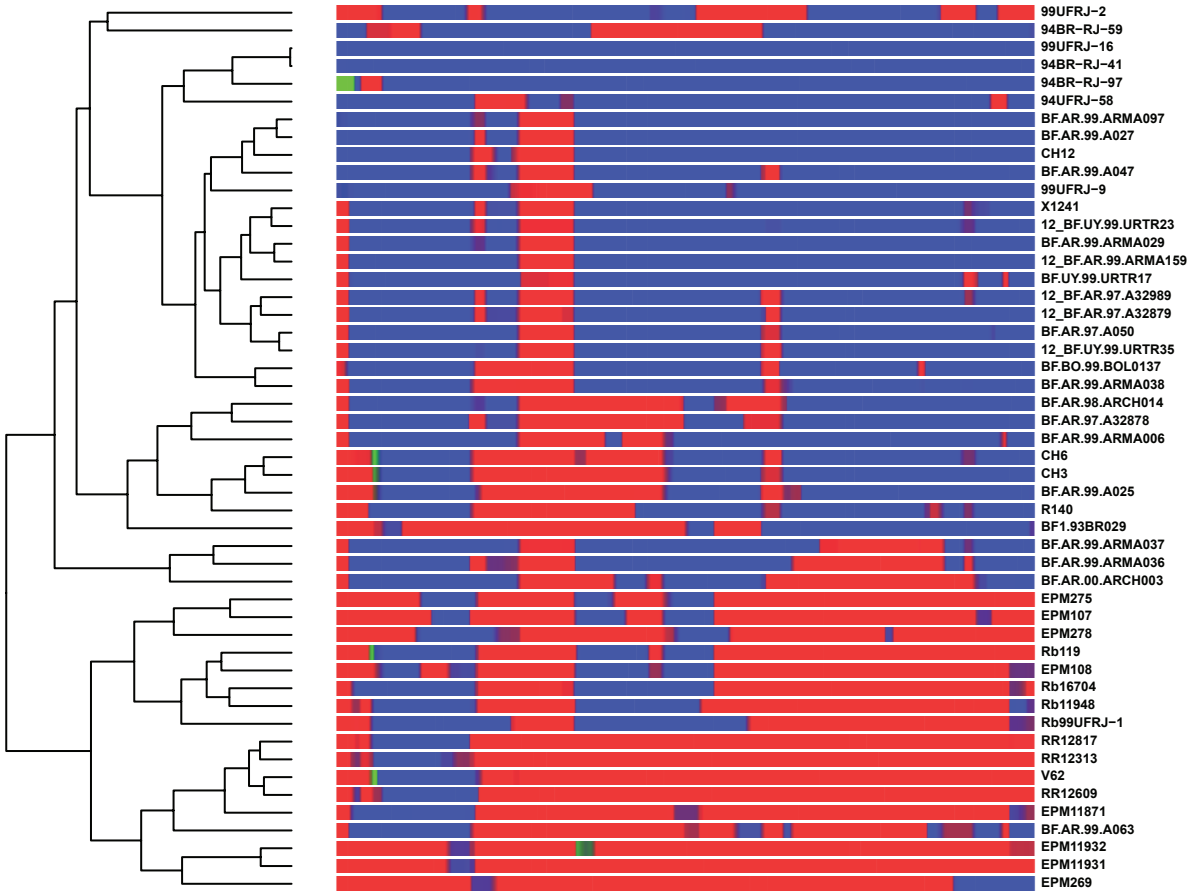

Figure S1. Mosaic composition of near full-length sequences of BF viruses isolated in South America. Genomic regions corresponding to subtype B viruses are depicted in red, regions corresponding to subtype F are depicted in blue and regions where the parental sequences could not be determined (low a posteriori probability) are depicted in green. The tree at left side of the figure represents the clustering pattern of South American isolates based on the genome breakpoints of these sequences.