Hindawi Publishing Corporation
AIDS Research and Treatment
Volume 2011, Article ID 154945, 13 pages
doi:10.1155/2011/154945

*Review Article*

# The Use of Bioinformatics for Studying HIV Evolutionary and Epidemiological History in South America

**Gonzalo Bello,[1,2] Marcelo A. Soares,[3,4] and Carlos G. Schrago[3]**

[1] *Laboratório de Imunologia e Aids, Instituto Oswaldo Cruz, 21040-360 Rio de Janeiro, RJ, Brazil*
[2] *Instituto de Microbiologia, Universidade Federal do Rio de Janeiro, 21941-590 Rio de Janeiro, RJ, Brazil*
[3] *Departamento de Genética, Universidade Federal do Rio de Janeiro, 21941-902 Rio de Janeiro, RJ, Brazil*
[4] *Programa de Genética, Instituto Nacional de Câncer, 20231-050 Rio de Janeiro, RJ, Brazil*

Correspondence should be addressed to Marcelo A. Soares, masoares@inca.gov.br

Received 27 May 2011; Accepted 19 August 2011

Academic Editor: Christina Ramirez Kitchen

Copyright © 2011 Gonzalo Bello et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The South American human immunodeficiency virus type 1 (HIV-1) epidemic is driven by several subtypes (B, C, and F1) and circulating and unique recombinant forms derived from those subtypes. Those variants are heterogeneously distributed around the continent in a country-specific manner. Despite some inconsistencies mainly derived from sampling biases and analytical constrains, most of studies carried out in the area agreed in pointing out specificities in the evolutionary dynamics of the circulating HIV-1 lineages. In this paper, we covered the theoretical basis, and the application of bioinformatics methods to reconstruct the HIV spatial-temporal dynamics, unveiling relevant information to understand the origin, geographical dissemination and the current molecular scenario of the HIV epidemic in the continent, particularly in the countries of Southern Cone.

## 1. Introduction

Human immunodeficiency virus (HIV), the causative agent of AIDS, is classified into types, groups, subtypes and subsubtypes according to its genetic diversity [1]. HIV type 1 (HIV-1) is widely disseminated worldwide and can be further divided into four genetic groups: group M (*major* or *main*), group O (*outlier*), and group N (*new* or *non-M non-O*), and the most recently characterized group P [1, 2]. While HIV-1 groups N, P, and O are restricted to countries of Central Africa, notably to Cameroon, HIV-1 group M is the responsible for the AIDS pandemic, accounting for over 90% of worldwide HIV infections [3]. HIV-2 is restricted to countries of West Africa, where it also represents a minority of viral infections and is decreasing in prevalence over time [4]. Nine pure subtypes of HIV-1 group M are currently known (A–D, F–H, J, and K). Some subtypes are further divided into subsubtypes, like subtypes F (F1 and F2) and A (A1, A2, and A3). Subtypes and sub-subtypes can form additional mosaic forms though recombination of different strains inside dually or multiply infected individuals [5]. Some of these recombinant forms may further achieve epidemic relevance, giving rise to known circulating recombinant forms (CRF). To date, at least 49 CRF are recognized in diverse parts of the world (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html).

It is currently accepted that HIV-1M subtypes and CRF are the result of founder effects in different geographic locales, followed by localized evolution. As a consequence, such HIV-1 forms are heterogeneously spread out worldwide [6]. Subtype B, for example, is primarily found in the Americas, Western Europe, Japan, and Australia. Subtype A is typical of some sub-Saharan African countries and Eastern European countries of the former Soviet Union. Subtype C is highly prevalent in countries of sub-Saharan Africa, India, and Brazil. Some CRF may also reach relevant epidemic status, and represent the predominant strains in certain geographic regions, such as CRF01_AE in Thailand and CRF02_AG in West African countries. Indeed, it has been recently suggested that CRF02_AG is the most rapid disseminating HIV-1 variant worldwide in the last years [6].

The differential distribution of HIV-1 subtypes and CRF also impact on their worldwide prevalence estimates. For instance, while subtype B dominates in several developed

countries with the lowest HIV prevalence rates, it accounts for only 11% of the worldwide infections. Conversely, subtype C accounts for nearly half of worldwide infections, as it prevails in countries with the highest HIV infection rates such as South Africa, and Botswana, or in highly populated countries like India [6].

## 2. HIV-1 Diversity in South America

South America follows the HIV molecular epidemiology commonly seen in the Americas, with HIV-1B being the most prevalent. However, a number of regional specificities are also observed. Brazil, the largest country of the continent, and which accounts for roughly two thirds of the infections, has likely the highest reported diversity. In addition to HIV-1B, other subtypes such as F1, C, and a number of B/F and B/C recombinants cocirculate [7–11]. In other South American countries, HIV-1B has been mostly reported, with the exception of Argentina and Uruguay, where a large number of B/F recombinants circulate at high proportion [12–16].

As HIV-1B, C, and F1 are the predominating pure subtypes in the area, a profusion of CRF comprising those subtypes have been characterized in South America. Those included the B/F-derived CRF12 in Argentina, Bolivia, Chile, Paraguay and Uruguay [17–19], CRF17 in Argentina and Paraguay [17, 19], CRF28, CRF29, CRF39, CRF40, and CRF46 in Brazil [20–22], CRF38 in Uruguay [15], CRF44 in Chile [23], and the B/C-derived CRF31 in southern Brazil [24]. All these CRF are thought to have been generated locally by recombination of the prevailing HIV-1 subtypes. Other HIV-1 clades have also been sporadically detected in South America, such as the subtype D and the CRF02_AG in Brazil [25–27] and the CFR16_AD and CRF06_cpx in Argentina [14, 28]. These forms were likely introduced by African immigrants, and have not achieved epidemic relevance to date.

## 3. Theoretical Basis of Evolutionary Bioinformatics Methods Applied on the Study of HIV

*3.1. Phylogenetics.* The most frequent evolutionary analysis performed on HIV sequences is usual phylogenetic inference. Phylogenetic trees provide essential information on the structure of the genetic diversity of the lineages. As in common hierarchical cluster analysis, phylogenies depict major groups in the data. However, such groups are related in time via the vertical process of genetic information passage, and, thus, tree topologies also depict the evolutionary history of the sequences. Phylogenies have been central to understand HIV evolution. The categorization of the major groups, subtypes, and their relationships were attained by phylogenetic inference [29]. Also, the characterization of HIV as a zoonosis and the identification of its geographic origins were permitted, because tree topologies for primate lentiviruses were known [30, 31].

Contemporaneously, phylogenetic tree reconstruction is accomplished by two different statistical approaches, maximum likelihood (ML) and Bayesian inference (BI). Both methods explicitly use Markov models of sequence evolution, and it is difficult to assert which one has superior performance [32]. ML inference would ideally address the phylogenetic problem by finding the tree that maximizes the likelihood function $f(D \mid T)$, where $D$ is the sequence alignment. In practice, however, such function is nonexistent because each topology has its own likelihood function $f(D \mid \theta)$, where $\theta$ is the vector of evolutionary parameters associated with a specific tree [33]. This peculiarity seems not to be an issue, since ML methods perform very well in simulations [34]. The Bayesian approach deals with the function $f(T \mid D)$, the posterior probability of the tree, which is fundamentally the product between the tree likelihood and the tree prior

$$f(T \mid D) = \frac{f(D \mid T) f(T)}{\sum_{i=1}^{S} f(D \mid T_i) f(T_i)}. \quad (1)$$

The denominator of the Bayes formula is the normalizing term. It is the sum of products between likelihood and prior of all $S$ tree topologies. Thus, the posterior probability of a given tree will lie between 0 and 1. This function is impossible to study analytically because the number of tree topologies conceivable is generally astronomical. Therefore, the posterior density is obtained via the Markov chain Monte Carlo (MCMC) technique [35].

Both ML and BI methods are computationally very intensive, but ML have become faster with recent developments of powerful heuristic search algorithms like PhyML [36], RAxML [37] and Garli [38], to cite a few. This has enabled the analysis of large data sets and the assessment of clade support via bootstrap [39]. BI is mainly implemented in the software MrBayes [40], which provides sophisticated models of sequence evolution and data handling. The use of MCMC algorithms, such as Metropolis-Hastings, bestows BI the capacity to adopt more complex models of sequence evolution in order to capture the biological reality of the evolutionary process [41].

In practice, if the researcher is interested in unveiling general patterns of genetic and spatial structuring of the viral diversity, phylogenetic inference should be performed on orthologous nonrecombining genome regions. This may be difficult to know *a priori*. Fortunately, there are analytical tools designed to identify recombination breakpoints on sequence alignments. For HIV, the SimPlot software [42] has been widely used. It implements the bootscanning, a simple sliding window strategy along the alignment in search of regions with conflicting phylogenetic signals [43]. Such regions will group with different reference sequences with significant bootstrap support. Although very useful and intuitive, this analysis does not offer a standard statistical testing framework. In this case, other methods have been recently developed. For instance, Pond and collaborators [44] have described a genetic algorithm to detect recombination breakpoints, implemented in GARD software (http://www.datamonkey.org/GARD/). This method uses the Akaike information criterion to choose among several breakpoints conformations.

*3.2. Phylogeography.* When studying the structure of the genetic diversity, it is often evident that lineages have a nonrandom distribution in space. In HIV epidemics, there are several examples of virus lineages of monophyletic origin with restricted spatial distribution [45]. When such a pattern is found, it means that the entrance of virus in the region was a unique event and it is possible to track its geographic origin by verifying the sister-group relationships on the phylogeny. Instead, if HIV sequences collected in an area are not monophyletic, we may still be able to track the geographic origins of the several independent virus lineages [46–49]. Actually, assessment of the spatial structure of the genetic diversity offers a relevant measure of the spatial dynamics of the epidemics that are critical to design public health policies [50].

Phylogeographic analysis has been gaining much attention from population biologists [51–53], since the spatial dynamics of organisms offer important insights into biological process from speciation to dispersion rates. Potentially applied to HIV research, a Bayesian implementation of ancestral reconstruction of spatial distribution was proposed by Lemey et al. [54], who implemented a Markov modeling of the discrete state of space of geographic localities, where the most parsimonious path is chosen by Bayesian stochastic search variable selection. This algorithm was subsequently extended to incorporate diffusion on a continuous space [55].

*3.3. Timescale.* Rooted phylogenetic trees impose a chronological direction on phylogenies. When rooted trees are ultrametric, branch lengths are proportional to the time elapsed since the separation of lineages. However, in the absence of external information on absolute times, chronologies can only be measured in units of mutations per site. If a calibration point is known, the branch lengths can be measured in absolute time (years, months, etc.). It is possible to extrapolate this information to the entire tree if evolutionary rates are homogeneous among lineages. This is the strict molecular clock [56]. In rapidly evolving pathogens, such as HIV, it is also possible to estimate divergence times by knowing the age of the leaves of the tree. This strategy is called tip dating and can be applied on heterochronous data sets, that is, when a significant number of mutations between sampling years occurs that enable the direct inference of the absolute mutation rate [57]. Populations that present such features are said to be measurably evolving [58].

As expected, the strict molecular clock rarely holds and, thus, a family of methods that estimates divergence times by relaxing the rate homogeneity assumption was developed over the last decade [59–63]. Although there are significant differences among these methods, they all share the same fundamental aim: the decomposition of branch lengths into absolute times and rates [64]. As in phylogenetic inference *per se*, ML and BI have been applied to tackle this problem. In an ML framework, branch lengths are decomposed by using multiple local molecular clocks [63, 65] or via rate-smoothing functions [66]. It is the Bayesian estimation of divergence time, however, that has gained much attention since its original proposition [35, 67]. This is mainly because the BI allows the usage of sophisticated models of evolutionary rate evolution, such as correlated [62] and uncorrelated models [59]. Moreover, calibration information can be flexibly incorporated by the adoption of probability distribution as priors [59].

In HIV research, ML and BI approaches have been widely applied in association with tip dating to estimate evolutionary rates and divergence times, both within and among hosts [68–71].

*3.4. Demography.* Another recent technical development in the study of HIV evolution is the application of methods derived from the coalescent theory. In a now classic paper, Kingman [72] derived the properties of genealogies obtained when population genetics is considered backwards in time. When doing this, several relevant parameters might be estimated, such as the time of the most recent common ancestor of alleles. The theoretical framework of the coalescent was extended to DNA sequences [73] and has also been applied to the study of evolutionary demography [74–76]. When the effective population size changes, the topology of the gene genealogy is expected to change in a predictable manner. Thus, demographic parameters like the growth rate might be inferred from the tree topology and different demographic models can be formally tested in a likelihood framework [75, 77]. For instance, it is possible to explicitly test if the likelihood of the data under the logistic growth model is significantly greater than the likelihood under the simple exponential growth or constant population size.

In a Bayesian framework, one may estimate the posterior probability of a demographic model given the data, $f(\Theta \mid D)$. Demographic models are incorporated via the coalescent prior function $f(g \mid \Theta)$, which computes the likelihood of a given genealogical tree topology $g$ given the demographic model $\Theta$. Evidently, the model of sequence evolution $\psi$ should also be considered, resulting in the following posterior distribution:

$$f(\Theta, \psi, g \mid D) = \frac{1}{Z} f(D \mid \psi, g) f(g \mid \Theta) f(\Theta) f(\psi), \quad (2)$$

where $Z$ is the denominator of the Bayes formula and $f(\Theta)$ and $f(\psi)$ are the demographic and sequence evolution model prior functions respectively. Thus, the marginal density $f(\Theta \mid D)$ is estimated by averaging over all $\psi$ and $g$ values

$$f(\Theta \mid D) = \frac{1}{Z} \sum_{g} \int_{\psi} f(D \mid \psi, g) f(g \mid \Theta) f(\Theta) f(\psi) d\psi. \quad (3)$$

From the above equation, it is evident that, in a Bayesian framework, tree topologies ($g$) are considered a nuisance parameter, since $\Theta$ is integrated over the topological space [78].

The demographic dynamics of HIV populations are, however, much more complex than the simplistic assumptions made by common growth models. Besides that, it is difficult to know *a priori* which model is appropriate to describe the demographic history. A family of coalescent techniques known as "skyline plots" was developed with the purpose of extracting demographic information from gene

genealogies without assuming an explicit model. Skyline plots depicts the variation of the effective population size through time by the adoption of a piecewise demographic mathematical description of the data [75]. As initially proposed, the method acts on a fixed ultrametric gene genealogy of $n$ terminals. For each $i$th interval between nodes, that is, coalescent events ($\Delta u$), the fundamental quantitative relation between the number of lineages ($k$) and effective population size ($\theta$) is applied via the function

$$\log f(g \mid \Theta, A) = \sum_{i=1}^{n-1} \log \frac{k_i(k_i-1)}{2\theta_i} - \frac{k_i(k_i-1)\Delta u_i}{2\theta_i}, \quad (4)$$

where $\Theta$ is the vector of effective population sizes and $A$ is number of classes used to group the $i$ intervals [74]. In the classic skyline plot, $A = i$. This equation describes some intuitive population genetics principles. For instance, if the effective population size decreases, the time interval between coalescent events will become shorter as the number of lineages increases.

Although this approach is much more realistic than the adoption of a specific model of population size change, there are still some drawbacks. Firstly, since the method acts on ultrametric trees, the strict molecular clock must be assumed. Also, the gene genealogy $g$ must be fixed, and hence, it is considered known without error. This is a twofold problem, because it is obvious that phylogenetic inference is subject to errors, and, frequently, the researcher is only interested in demographic parameters instead of topological relationships of the sequences. The proposition of the Bayesian skyline down sized these problems [74]. In a Bayesian context, gene genealogies $g$ may be considered a nuisance parameter and demographic parameters can be estimated by integrating over the topological space. This is achieved via MCMC algorithms [78]. The Bayesian skyline method needs *a priori* determination of the number $A$ of intervals between coalescent events. This is subjective and largely depends on the historical information content of the alignment. Another weakness is that effective population sizes are assumed to be correlated in successive coalescent intervals. To surpass this issue, the Bayesian skyride, a method that penalizes the change of this parameter between intervals was developed [79].

Piecewise demographic models are recommended to be applied when sequence alignments bear significant demographic information [80]. In practice, this is difficult to determine, but studies involving measurably evolving populations, like the majority of HIV datasets, are suitable for such analysis. Ideally, the power of the skyline methods is increased when multiple *loci* are used. In order to incorporate the information from multiple *loci*, Heled and Drummond [81] have proposed the extended skyline method. However, to gain statistical power, multiple unlinked nonrecombining *loci* must be used. Unfortunately, this is not feasible for HIV datasets and researchers may only try to reduce stochastic error of the analysis by augmenting the number of nucleotide sites. For the moment, none of these methods have incorporated population structure in their framework. Since it is not clear how spatial population structure affects skyline plots, when such information is known *a priori*, it is better to investigate each population separately.

Finally, the coalescent theory used in demographic estimation measures time in generations. Therefore, when calibrating a gene genealogy, in which branch lengths are measured as the number of mutations per site, one should ideally enter chronological information in generations ($\tau$). When this is the case, skyline plots will depict the variation of the absolute effective population size ($N_e$) through the generations. Most commonly though, in HIV studies, chronological information is measured in years via tip dating. Thus, the unit of the $y$-axis of the skyline plots is the product $N_e \times \tau$, where $\tau$ is the generation time in years.

Demographic inference using the methods described above is basically implemented on the BEAST software [80]. In practice, researchers will use the skyline model as topological prior while simultaneously inferring divergence times and evolutionary rates in a relaxed or strict clock framework. Therefore, in a single analysis, population demography, the time of the most recent common ancestor of lineages and evolutionary rates are coestimated considering tree topology as a nuisance parameter, since values are averaged over the topologies sampled during the MCMC run. Actually, these tree topologies might be summarized to graphically represent historical process that generated the sequences.

*3.5. Origin and Timescale of HIV-1 Clades in South America.* Several studies have been performed to reconstruct the origin and timescale of major HIV-1 clades circulating in South America, including subtypes B, C, F1, and several CRF lineages.

Subtype B viruses circulating in South America belong to the "pandemic" clade that migrated out of Haiti around 1969 (1966–1972) and spread through the world [82]. In most of the South American countries, HIV-1B epidemic probably resulted from introduction of multiple strains and subsequently spread within local networks although this hypothesis has not been formally tested. Some country-specific subtype B polymorphisms, however, have been described in South America. While most (~95%) subtype B viruses of the pandemic clade carry a GPGR motif at the tip of the V3 loop, the Brazilian subtype B epidemic is characterized by roughly similar proportions of strains containing the common GPGR motif and the unusual GWGR motif [9, 83–86]. Phylogenetic analyses of Brazilian subtype B *env* sequences showed that GWGR isolates formed a monophyletic cluster (B-Br clade) nested within the basal GPGR Brazilian sequences [87, 88], supporting the hypothesis that GWGR strains originated from a single founder GPGR Brazilian strain. The TMRCA of the B-Br clade was estimated to be 1966 (1954–1975) [87], which roughly coincides with the age of the subtype B pandemic clade, suggesting that the founder event that originates the Br-B lineage probably occurred at the beginning of the subtype B epidemic in Brazil. This hypothesis is supported by a recent analysis of HIV-positive serum samples collected in Brazil in 1983 that confirms the circulation of GWGR isolates at such very early stage of the Brazilian epidemic [89]. Recent phylogenetic studies using full-length subtype B genomes

showed that GWGR viruses are evenly dispersed among GPGR Brazilian strains [84, 89], suggesting a polyphyletic origin of GWGR strains. Such observation, however, could result from extensive intrasubtype recombination between GWGR and GPGR variants, rather than from independent evolution of multiple GPGR strains into GWGR strains.

The circulation of "pure" subtype F1 viruses in South America seems to be almost restricted to Brazil. Despite the high prevalence of BF1 recombinants in countries from the Southern cone (Argentina, Brazil, Bolivia, Chile, Paraguay, and Uruguay), full-length subtype F1 viruses, or even subtype F1 *pol* sequences, are very rarely found outside Brazil [13–16, 18, 90]. Phylogenetic analyses of full-length and partial genome subtype F1 sequences consistently showed that Brazilian and South American viruses form a monophyletic group when compared to subtype F1 viruses from other countries around the world [87, 90–93]. Within such South American lineage, subtype F1 *env* fragments of BF1 recombinants from Argentina, Bolivia, Chile, Paraguay, and Uruguay form a monophyletic cluster nested within the basal subtype F1 Brazilian sequences [90]. These results indicate that South American HIV-1F1 and HIV-1BF1 epidemics are the result of the introduction of a single founder subtype F1 strain through Brazil, followed by expansion and recombination of this virus with local subtype B viruses. Recent evidence suggests that this founder strain came from the Democratic Republic of Congo (DRC), as it does not resemble other HIV-1F1 worldwide lineages such as those found in Angola and Romania [91]. Three independent studies based on the analysis of *env* and *pol* sequences date back the TMRCA of the Brazilian and South American subtype F1 clade to between the middle 1970s and the early 1980s: 1976 (1966–1982) [90], 1978 (1972–1983) [87], and 1980 (1975–1985) [93]. One study based on the analysis of *gag* sequences, however, traced the TMRCA of South American subtype F1 clade back to 1969 (1959–1978) [92], similar to the origin of the subtype B pandemic clade.

The pervasive recombination between subtype B and F1 viruses in South America created a large variety of intersubtype BF1 recombinants, some of which have disseminated across several individuals, gaining the status of CRFs_BF. The most widespread of these CRFs is the CRF12_BF that circulates in Argentina, Uruguay, Bolivia, Chile, and Paraguay. Epidemiological data revealed that CRF12-like BF1 viruses have been circulating in Argentina since the mid 1980s [94], but the exact origin of this recombinant clade is still uncertain. Three studies have reconstructed the timescale of the CRF12_BF using Bayesian relaxed-clock methods, with quite different results. The first study, based on the analysis of *vpu* CRF12-like sequences from Argentine children, estimated the TMRCA of this clade at 1992 (1981–1996) [95]. The second study, based on the analysis of a large data set of *pol* CRF12-like sequences from Argentina and Uruguay, dated back the origin of this clade to 1983 (1978–1988) [96]. The third study, based on the analysis of subtype B *pol* gene fragments from CRF12_BF viruses from Argentina, suggests that the origin of this CRF could be dated back to 1969 although the confidence interval (CI) of such estimate was extremely large (1946–1981) [92]. The timescale of other CRF_BF viruses with more restricted circulation have been also estimated. The TMRCA of the CRF38_BF clade that circulates in Uruguay was traced to 1986 (1981–1990) [96], while the CRF28_BF and CRF29_BF clades circulating in the southeastern region of Brazil probably evolved from a common BF1 recombinant ancestor that existed around 1989 (1987–1993) [97]. Thus, the South American CRFs_BF represent "old" viral lineages probably generated during the 1980s, shortly after the introduction of subtype F1 into the region.

The circulation of HIV-1C in South America is mainly concentrated in the southern region of Brazil. Most studies performed to date support the notion that South American HIV-1C epidemic was also the result of a single founder event followed by local dissemination of the new virus and suggested the entrance spot at southern Brazil [11, 98–101]. Although one study claimed a possible entrance of HIV-1C through Argentina [102], epidemiological data are not consonant with this hypothesis. Several studies have been performed to trace the origin of HIV-1C lineage that colonized the region. Some studies traced the origin of such clade to somewhere in East Africa, most likely to Burundi, Kenya, or Ethiopia [98, 99]. Others proposed the introduction of HIV-1C in southern Brazil from Mozambique by means of Portuguese colonization of the latter country and migration to Brazil [103], but phylogenetic and molecular evidence did not support that proposal [104]. An additional study suggests that HIV-1C migrated from East Africa to Brazil through a network of men who have sex with men (MSM) from London, England [100]. More detailed analyses and additional samples are, however, required to fully elucidate such relationships and the actual migration history of subtype C to Brazil. It is also unclear when the HIV-1C founder strain was introduced into the Brazilian population. The first study to estimate the timescale of Brazilian subtype C lineage used an ML strict-clock approach and points the origin of such clade to the early 1990s [105]. A second study employed a Bayesian strict-clock method and estimates the onset date of the Brazilian subtype C epidemic at around 1987, but the CI obtained was very wide (1956–1998) [106]. More recent studies based on Bayesian relaxed-molecular clock models described older mean TMRCA estimates. Two independent studies indicate that the TMRCA of Brazilian subtype C clade dates back to the early 1980s: 1980 (1972–1987) [100] and 1982 (1972–1988) [98]. Another study suggests that subtype C introduction in Brazil could be even older, dating back to between 1960 and 1970, although important variations were observed in such study across distinct viral genomic regions analyzed: gp41 (1962; 1950–1972), RT (1968; 1959–1976), and p24 (1977; 1966–1986) (Figure 1) [101].

The cocirculation of HIV-1 subtypes B and C in the southern Brazilian region also creates a variety of intersubtype BC recombinants, including one CRF designated as CRF31_BC, which is particularly prevalent in Rio Grande do Sul, the southernmost state of Brazil. Phylogenetic and informative site analyses comparing the CRF31_BC lineage with Brazilian subtypes B and C clades clearly supports a local origin of this recombinant form [24, 107]. Two

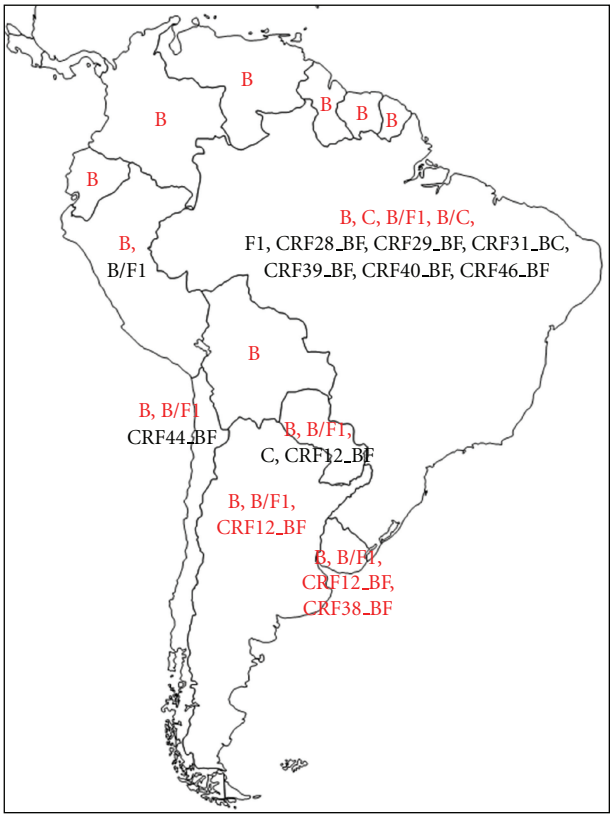



Figure 1: Political map of South America depicting the major HIV-1 variants occurring in each country. Major variants (prevalence above 5%) are listed in red.

independent studies have employed Bayesian clock methods to estimate the timescale of the CRF31_BC epidemic using the recombinant *pol* (PR/RT) gene fragment. Both studies estimated the TMRCA of the CRF31_BC clade at around the late 1980s: 1987 (1967–1998) [106] and 1988 (1979–1993) [107]. The identification of CRF31_BC-infected Brazilian individuals with HIV diagnosis as early as 1990 [106, 107] is fully consistent with the estimated origin of this clade during the 1980s. Although CRF31_BC viruses certainly derived from a single BC recombinant ancestor, phylogenetic analyses of subtype C genomic regions (integrase, *env*-gp120 and *env*-gp41) from CRF31_BC viruses reveal that those viruses do not formed a monophyletic cluster within subtype C Brazilian clade, but were evenly dispersed among subtype C Brazilian viruses [107]. This observation resembles the lack of monophyletic clustering of subtype B GWGR variants outside the *env* region and could be also explained by the widespread recombination between CRF31_BC and subtype C Brazilian strains.

The different studies performed up to date support two opposite scenarios for the timescale of the HIV-1 epidemic in South America. Some studies suggest that HIV-1B was the first to colonize the continent between the middle 1960s and the early 1970s, followed by HIV-1F1 and HIV-1C some years later (between the middle 1970s and the early 1980s). Other studies, however, supports the concurrent introduction of all three HIV-1 subtypes in South American between the middle 1960s and the early 1970s. Determine the actual timescale of the major HIV-1 clades circulating in South America is of paramount importance for understanding the circumstances surrounding the emergence of such epidemics and their subsequent dissemination dynamics.

The great variation in the mean estimated TMRCA of some South American clades: subtype C (1962 to 1990), subtype F1 (1969 to 1980), and CRF12_BF (1969 to 1992), as well as the wide range of the CI of many estimates (up to 30 years) exposes the important challenge to derive reliable timescales for HIV evolution. The considerable rate variation among HIV-1 lineages at the population level produces a departure from the clock-like evolution that can seriously hamper our ability to accurately estimate the evolutionary rate and the TMRCA of HIV-1 [108]. However, uncertainty in TMRCA estimates of South American HIV clades remains despite the implementation of more realistic and phylogenetically accurate "relaxed" molecular clocks models that accommodate such rate variation among lineages [59]. It is possible that variation in the size and the nature of data sets [109], and/or time intervals at which sequences are sampled [110] may also disturb the substitution rates and divergence date estimations.

*3.6. Demographic History and Epidemic Potential of HIV-1 Clades in South America.* The development of phylogenetic approaches that incorporate the coalescent theory of population genetics enabled us to infer the demographic history and epidemic potential of major South American HIV-1 clades.

The first study published in 2005 used an ML coalescent-based approach to explore the demographic history and epidemic potential of HIV-1 subtypes B and C in Brazil, based on the analysis of *pol* sequences collected up to 2001 [105]. That study suggested that both HIV-1 subtypes were spreading exponentially in Brazil and that the mean subtype C growth rate (0.6–0.8 year$^{-1}$) was about twice that of subtype B (0.2–0.4 year$^{-1}$) (Table 1). These observations were confirmed by a second study that used a Bayesian coalescent-based approach to estimate the growth rate of Brazilian subtypes B and C epidemics under a demographic model of exponential growth (Table 1) [106]. Thus, initial studies supported the existence of a growing HIV-1 epidemic in Brazil by the 2000s and suggest that subtype C was spreading at a significantly faster rate than subtype B.

Subsequent studies, however, pointed to a different epidemic scenario. Bayesian skyline plot analyses of HIV-1 *env* and *pol* sequences collected from Brazilian patients up to 2005–2006 indicated that subtype B and F1 epidemics in the southeastern region and subtype C epidemic in the southern region were better explained by a model of logistic growth, characterized by an initial period of rapid exponential expansion followed by a decline in growth rate since 1985–1995 [87, 111]. Such proposed slowdown of the growth rate of HIV-1 subtypes B, F1 and C epidemics coincides with epidemiological information that reveals that after a period of explosive growth during the 1980s and 1990s, the number of new AIDS cases annually reported in the southeastern and

Table 1: Coalescent estimates of epidemic growth rate of HIV-1 clades in South America.

| Subtype | Demographic model | Molecular clock | Gene | Growth rate(year$^{-1}$) | Reference |
|---|---|---|---|---|---|
| B | Exponential growth | Strict | *pol* (PR) | 0.23 (0.20–0.26) | [105] |
| | | | *pol* (RT) | 0.35 (0.31–0.40) | |
| | | | *pol* (PR/RT) | 0.4 (0.2–0.7) | [106] |
| | Logistic growth | Strict | *env (C2-V3)* | 0.54 (0.32–0.78) | [87] |
| | | | *pol (PR-RT)* | 0.56 (0.35–0.80) | |
| F1 | Logistic growth | Strict | *env (C2-V3)* | 0.61 (0.40–0.86) | [87] |
| | | | *pol (PR-RT)* | 0.59 (0.31–0.92) | |
| C | Exponential growth | Strict | *pol* (PR) | 0.77 (0.62–0.93) | [105] |
| | | | *pol* (RT) | 0.63 (0.51–0.75) | |
| | | | *pol* (PR/RT) | 0.7 (0.3–1.0) | [106] |
| | Logistic growth | Strict | *env (C2-V3)* | 0.77 (0.45–1.14) | [111] |
| | | | *pol (RT)* | 0.70 (0.41–1.00) | |
| | | Relaxed | *env (C2-V3)* | 0.87 (0.50–1.29) | |
| | | | *pol (RT)* | 0.81 (0.40–1.26) | |
| CRF12_BF | Logistic growth | Relaxed | *vpu* | 2.24 (0.21–4.56) | [95] |
| | | Strict | *pol (PR-RT)* | 1.08 (0.79–1.44) | [96] |
| | | Relaxed | | 1.22 (0.85–1.64) | |
| CRF28_BF/CRF29_BF | Logistic growth | Relaxed | *pol (PR-RT)* | 1.18 (0.64–1.38) | [97] |
| | | | *gag* | 1.20 (0.59–1.47) | |
| CRF31_BC | Logistic growth | Strict | *pol (RT)* | 1.26 (0.61–2.10) | [111] |
| | | Relaxed | | 1.27 (0.44–2.26) | |
| CRF38_BF | Logistic growth | Strict | *pol (PR-RT)* | 0.83 (0.31–1.81) | [96] |
| | | Relaxed | | 0.92 (0.41–1.75) | |

southern Brazilian regions has shown a trend toward stability since 1995 and 2000, respectively [112].

The mean growth rate of Brazilian subtype B epidemic estimated for the model of logistic growth (0.45–0.55 year$^{-1}$) [87] was significantly higher than that previously estimated for the model of exponential growth [105], yet lower than that described for the North American subtype B epidemic (0.8 year$^{-1}$) under the same logistic demographic pattern [113]. By contrast, the mean estimated growth rate of subtype C epidemic for the model of logistic growth (0.7–0.9 year$^{-1}$) [111] was similar to that previously obtained for the exponential one [105]. The mean initial growth rate of subtype F1 epidemic (~0.6 year$^{-1}$) was in-between those estimated for subtype B and C [87]. Although the mean

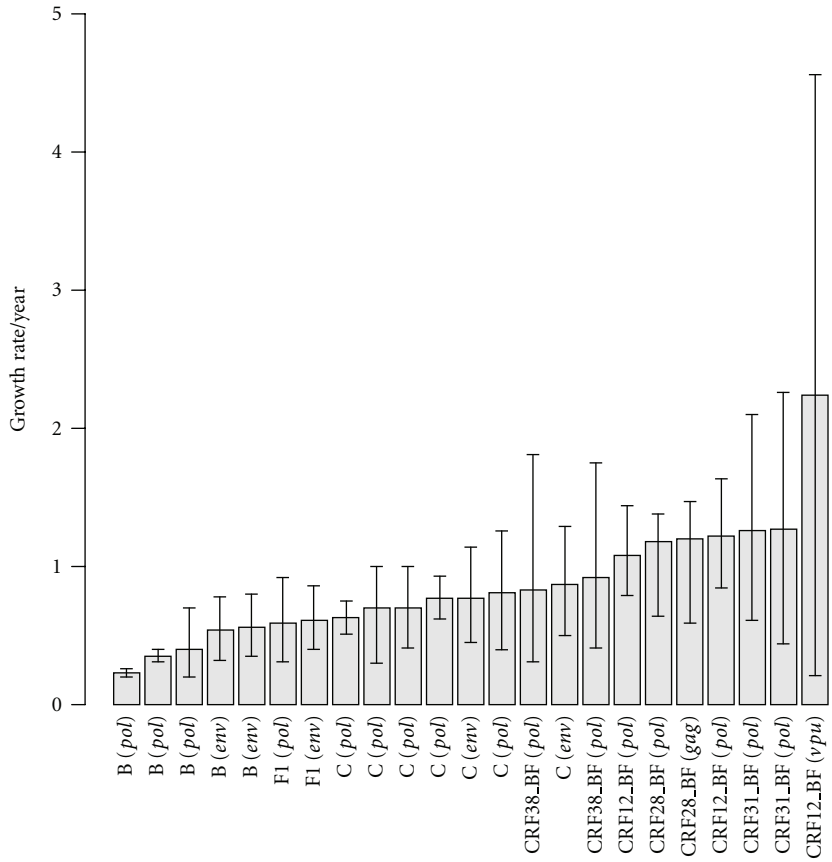



Figure 2: Histogram depicting the variance in the epidemic growth rates of HIV-1 clades in South America. Refer to Table 1 for further details regarding applied models and references describing estimates.

growth rates for the logistic growth model support the notion that subtype C clade exhibited an initial rate of spread slightly higher than subtypes B and F1, the CI intervals of such estimates displayed a great overlap (Table 1 and Figure 2). Thus, it is unclear whether the initial rate of dissemination of different Brazilian HIV-1 subtypes was or not significantly different.

Some studies have also used the Bayesian skyline coalescent-based method to reconstruct the epidemic history of major South American CRFs clades. All studies indicate that the effective number of infections by CRF_BF (12, 28/29, and 38) and CRF31_BC experienced a fast exponential growth over a 5–15 year period after their emergence, but then decreased toward the present following the same logistic model of population growth described for parental HIV-1 subtypes [95–97, 111]. An initial study of the CRF12_BF epidemic in Argentine children indicated an extremely rapid rate of population expansion for this clade (2.2 year$^{-1}$), but the CI of such estimate was huge (0.21–4.56 year$^{-1}$) [95]. A more precise estimate was recently obtained through the analysis of CRF12_BF viruses circulating in Argentina and Uruguay [96]. According to this study, the CRF12_BF epidemic spread in those countries with an initial mean growth rate of around 1.2 year$^{-1}$ (0.85–1.6 year$^{-1}$), which is about half of that previously obtained for this CRF, but still higher than those reported for Brazilian HIV-1 subtypes B, F1, and C. Similarly, high initial mean growth rates were also recently estimated for the CRF38_BF clade in Uruguay (0.9 year$^{-1}$) [96], and for the CRF28/29_BF (1.2 year$^{-1}$) [97] and CRF31_BC (1.3 year$^{-1}$) [111] clades in Brazil (Table 1).

Thus, Bayesian coalescent-based analyses performed to date suggest that major HIV-1 clades circulating in South America followed the same overall demographic pattern described for HIV-1 subtype B in USA and some European countries [46, 113–116], characterized by an initial phase of rapid expansion followed by a recent period of stabilization. Such a recent decline in the growth rate of these HIV-1 epidemics may be the consequence of implementation of efficient prevention campaigns after the official recognition of HIV/AIDS in the early 1980s, and/or the result of a saturation of high-risk transmission networks in concentrated HIV/AIDS epidemics. One important limitation of the studies performed in Brazil, however, is that most HIV-1 samples were derived from the major metropolitan areas of the southern and southeastern regions, and may not represent the demographic trend of HIV epidemics in other localities. It has been documented that while the number of new AIDS cases has remained stable over the last years in the cities with >500,000 inhabitants from the southern, southeastern, and central-west regions, that number continuous to growth in

the northern and northeastern regions as well as in the cities with <50,000 inhabitants from all over the country [112].

Despite some overlap of the CI of growth rates estimates (Table 1), Bayesian coalescent-based analyses also suggest that CRFs have spread at a rate much higher than parental HIV-1 subtypes in the South American population (Figure 2). A very attractive hypothesis to explain such observation is to propose that CRFs display a higher transmissibility than their parental HIV-1 subtypes. Indeed, such scenario has been suggested to explain the predominance of CRF02_AG in West Central Africa [117]. According to this hypothesis, one should expect that CRFs may eventually become prevalent in the entire region over time. The prevalence of CRFs, however, displays a great variation across neighboring regions in South America. While CRF12_BF (and related BF recombinants) attains a very high prevalence (>50%) in Argentina [12–14] and Uruguay [15], it displays an intermediate prevalence (<20%) in Chile and Paraguay [18, 19] and is almost completely absent in Brazil [11, 118–120]. Moreover, there is recent evidence that this CRF may be declining in prevalence in Argentina, through the analysis of vertically infected children [121]. Similarly, although the CRF28/29_BF and CRF31_BC variants reach a high prevalence in the cities of Santos (Sao Paulo state) [122] and Porto Alegre (Rio Grande do Sul state) in Brazil [106, 123], respectively, they are rarely found in other neighboring cities [120, 123, 124]. This indicates a fast, but geographically contained expansion of those CRFs, sometimes limited to a specific locality.

An alternative hypothesis suggests that difference in the rate of expansion of distinct HIV-1 clades in South America may reflect a variation in the efficiency of different transmission networks. According to this hypothesis, some HIV-1 clades spread faster, because they encounter a more favorable local transmission chain. It has been proposed that the prevalence of CRF12_BF is higher in South American countries whit more extensive IDU epidemics [16]. IDU populations are thought to represent extremely fast chains of virus transmission, and initial expansion of CRF12_BF through such efficient networks may explain its rapid initial growth rate in Argentina and Uruguay. Of note, a fast rate of dissemination (1.5 $\text{years}^{-1}$) was reported for HIV-1B in a cohort of men having sex with men (MSM) in Italy [116], while variable growth rates (from 0.5 $\text{years}^{-1}$ to 1.4 $\text{years}^{-1}$) were described for HIV-1B spreading in a number of MSM transmission chains in the UK [114]. These evidences support the notion that the rate of expansion of HIV-1 in a given population is determined by the efficiency of the transmission network, rather than by the specific genetic composition of the viral strain.

The history of HIV epidemic in South America has been largely clarified by the application of bioinformatic tools developed from evolutionary genetics models of demography, population genetics, and phylogenetics. Over the last decades, novel algorithms have been implemented that improved parameter estimates and unveiled unique aspects of HIV spatial-temporal dynamics in the South American continent. However, unresolved issues still remain regarding uneven geographical and chronological sampling that warrant further assessment, which once overcome will greatly enhance the robustness of the estimates.

## Acknowledgments

All three authors are supported by the Brazilian Ministry of Education through their Professorship status at Universidade Federal do Rio de Janeiro. G. Bello and M. A. Soares are also supported by the Brazilian Ministry of Health. Additional funding is from the Brazilian Research Council-CNPq grants no. 304416/2010-0 and 308147/2009-0 (to M. M. A. Soares and C. G. Schrago, resp.), and from the Rio de Janeiro State Science Foundation-FAPERJ grants no. 102.858/2008 to M. A. Soares and 103.136/2008, 110.838/2010 and 110.028/2011 to C. G. Schrago.

## References

[1] D. L. Robertson, J. P. Anderson, J. A. Bradac et al., "HIV-1 nomenclature proposal," *Science*, vol. 288, no. 5463, pp. 55–57, 2000.

[2] J. C. Plantier, M. Leoz, J. E. Dickerson et al., "A new human immunodeficiency virus derived from gorillas," *Nature Medicine*, vol. 15, no. 8, pp. 871–872, 2009.

[3] D. M. Tebit and E. J. Arts, "Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease," *The Lancet Infectious Diseases*, vol. 11, no. 1, pp. 45–56, 2011.

[4] S. Eholié and X. Anglaret, "Commentary: decline of HIV-2 prevalence in West Africa: good news or bad news?" *International Journal of Epidemiology*, vol. 35, no. 5, pp. 1329–1330, 2006.

[5] D. S. Burke, "Recombination in HIV: an important viral evolutionary strategy," *Emerging Infectious Diseases*, vol. 3, no. 3, pp. 253–259, 1997.

[6] J. Hemelaar, E. Gouws, P. D. Ghys, and S. Osmanov, "Global trends in molecular epidemiology of HIV-1 during 2000–2007," *AIDS*, vol. 25, no. 5, pp. 679–689, 2011.

[7] V. Bongertz, D. C. Bou-Habib, L. F. M. Brígido et al., "HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites," *Journal of Acquired Immune Deficiency Syndromes*, vol. 23, no. 2, pp. 184–193, 2000.

[8] J. C. Couto-Fernandez, M. G. Morgado, V. Bongertz et al., "HIV-1 subtyping in Salvador, Bahia, Brazil: a city with African sociodemographic characteristics," *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, vol. 22, no. 3, pp. 288–293, 1999.

[9] M. G. Morgado, E. C. Sabino, E. G. Shpaer et al., "V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F," *AIDS Research and Human Retroviruses*, vol. 10, no. 5, pp. 569–576, 1994.

[10] E. C. Sabino, E. G. Shpaer, M. G. Morgado et al., "Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil," *Journal of Virology*, vol. 68, no. 10, pp. 6340–6346, 1994.

[11] M. A. Soares, T. de Oliveira, R. M. Brindeiro et al., "A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil," *AIDS*, vol. 17, no. 1, pp. 11–21, 2003.

[12] M. M. Thomson, M. L. Villahermosa, E. Vázquez-De-Parga et al., "Widespread circulation of a B/F intersubtype recombinant form among HIV-1-infected individuals in Buenos Aires, Argentina," *AIDS*, vol. 14, no. 7, pp. 897–899, 2000.

[13] J. F. Quarleri, A. Rubio, M. Carobene et al., "HIV type 1 BF recombinant strains exhibit different pol gene mosaic patternsdscriptive analysis from 284 patients under treatment failure," *AIDS Research and Human Retroviruses*, vol. 20, no. 10, pp. 1100–1107, 2004.

[14] D. A. Dilernia, A. M. Gomez, L. Lourtau et al., "HIV type 1 genetic diversity surveillance among newly diagnosed individuals from 2003 to 2005 in Buenos Aires, Argentina," *AIDS Research and Human Retroviruses*, vol. 23, no. 10, pp. 1201–1207, 2007.

[15] D. Ruchansky, C. Casado, J. C. Russi, J. R. Arbiza, and C. Lopez-Galindez, "Identification of a new HIV Type 1 circulating recombinant form (CRF38-BF1) in Uruguay," *AIDS Research and Human Retroviruses*, vol. 25, no. 3, pp. 351–356, 2009.

[16] J. Hierholzer, S. Montano, M. Hoelscher et al., "Molecular epidemiology of HIV type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina," *AIDS Research and Human Retroviruses*, vol. 18, no. 18, pp. 1339–1350, 2002.

[17] J. K. Carr, M. Avila, M. G. Carrillo et al., "Diverse BF recombinants have spread widely since the introduction of hiv-1 into South America," *AIDS*, vol. 15, no. 15, pp. F41–F47, 2001.

[18] M. Ríos, E. Belgado, L. Pérez-Álvarez et al., "Antiretroviral drug resistance and phylogenetic diversity of HIV-1 in Chile," *Journal of Medical Virology*, vol. 79, no. 6, pp. 647–656, 2007.

[19] N. Aguayo, V. A. Laguna-Torres, M. Villafane et al., "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," *Revista da Sociedade Brasileira de Medicina Tropical*, vol. 41, no. 3, pp. 225–231, 2008.

[20] D. J. De Sá Filho, M. C. A. Sucupira, M. M. Casiero, E. C. Sabino, R. S. Diaz, and L. M. Janini, "Identification of two HIV type 1 circulating recombinant forms in Brazil," *AIDS Research and Human Retroviruses*, vol. 22, no. 1, pp. 1–13, 2006.

[21] M. L. Guimarães, W. A. Eyer-Silva, J. C. Couto-Fernandez, and M. G. Morgado, "Identification of two new CRF_BF in Rio de Janeiro State, Brazil," *AIDS*, vol. 22, no. 3, pp. 433–435, 2008.

[22] S. S. Sanabani, E. R. De Souza Pastena, W. K. Neto, V. P. Martinez, and E. C. Sabino, "Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil," *Virology Journal*, vol. 7, article 74, 2010.

[23] E. Delgado, M. Ríos, J. Fernández, L. Pérez-Álvarez, R. Nájera, and M. M. Thomson, "Identification of a new HIV type 1 BF intersubtype circulating recombinant form (CRF44-BF) in Chile," *AIDS Research and Human Retroviruses*, vol. 26, no. 7, pp. 821–826, 2010.

[24] A. F. Santos, T. M. Sousa, E. A. J. M. Soares et al., "Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil," *AIDS*, vol. 20, no. 16, pp. 2011–2019, 2006.

[25] W. A. Eyer-Silva and M. G. Morgado, "Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil," *Memorias do Instituto Oswaldo Cruz*, vol. 102, no. 7, pp. 809–815, 2007.

[26] L. F. A. MacHado, M. O. G. Ishak, A. C. R. Vallinoto et al., "Molecular epidemiology of HIV type 1 in Northern Brazil: identification of subtypes C and D and the introduction of CRF02-AG in the amazon region of Brazil," *AIDS Research and Human Retroviruses*, vol. 25, no. 10, pp. 961–966, 2009.

[27] J. C. Couto-Fernandez, W. A. Eyer-Silva, M. L. Guimarães et al., "Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil," *AIDS Research and Human Retroviruses*, vol. 22, no. 2, pp. 207–211, 2006.

[28] M. Gómez-Carrillo, J. F. Quarleri, A. E. Rubio et al., "Drug resistance testing provides evidence of the globalization of HIV type 1: a new circulating recombinant form," *AIDS Research and Human Retroviruses*, vol. 20, no. 8, pp. 885–888, 2004.

[29] A. Rambaut, D. Posada, K. A. Crandall, and E. C. Holmes, "The causes and consequences of HIV evolution," *Nature Reviews Genetics*, vol. 5, no. 1, pp. 52–61, 2004.

[30] B. H. Hahn, G. M. Shaw, K. M. De Cock, and P. M. Sharp, "AIDS as a zoonosis: scientific and public health implications," *Science*, vol. 287, no. 5453, pp. 607–614, 2000.

[31] B. F. Keele, F. Van Heuverswyn, Y. Li et al., "Chimpanzee reservoirs of pandemic and nonpandemic HIV-1," *Science*, vol. 313, no. 5786, pp. 523–526, 2006.

[32] J. P. Huelsenbeck, B. Larget, R. E. Miller, and F. Ronquist, "Potential applications and pitfalls of Bayesian inference of phylogeny," *Systematic Biology*, vol. 51, no. 5, pp. 673–688, 2002.

[33] Z. Yang, "Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods," *Systematic Biology*, vol. 43, no. 3, pp. 329–342, 1994.

[34] J. P. Huelsenbeck, "Performance of phylogenetic methods in simulation," *Systematic Biology*, vol. 44, no. 1, pp. 17–48, 1995.

[35] B. Mau, M. A. Newton, and B. Larget, "Bayesian phylogenetic inference via Markov chain Monte Carlo methods," *Biometrics*, vol. 55, no. 1, pp. 1–12, 1999.

[36] S. Guindon and O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Systematic Biology*, vol. 52, no. 5, pp. 696–704, 2003.

[37] A. Stamatakis, T. Ludwig, and H. Meier, "RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees," *Bioinformatics*, vol. 21, no. 4, pp. 456–463, 2005.

[38] M. J. Brauer, M. T. Holder, L. A. Dries, D. J. Zwickl, P. O. Lewis, and D. M. Hillis, "Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference," *Molecular Biology and Evolution*, vol. 19, no. 10, pp. 1717–1726, 2002.

[39] J. Felsenstein, "Confidence-limits on phylogenies with a molecular clock," *Systematic Zoology*, vol. 34, pp. 152–161, 1985.

[40] F. Ronquist and J. P. Huelsenbeck, "MrBayes 3: Bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, 2003.

[41] J. P. Huelsenbeck, F. Ronquist, R. Nielsen, and J. P. Bollback, "Bayesian inference of phylogeny and its impact on evolutionary biology," *Science*, vol. 294, no. 5550, pp. 2310–2314, 2001.

[42] Ray S. Simplot v2.5.0, http://www.hopkinsmedicine.org/.

[43] M. O. Salminen, J. K. Carr, D. S. Burke, and F. E. McCutchan, "Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning," *AIDS Research and Human Retroviruses*, vol. 11, no. 11, pp. 1423–1425, 1995.

[44] S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost, "GARD: a genetic algorithm for recombination detection," *Bioinformatics*, vol. 22, no. 24, pp. 3096–3098, 2006.

[45] M. M. Thomson and R. Nájera, "Molecular epidemiology of HIV-1 variants in the global aids pandemic: an update," *AIDS Reviews*, vol. 7, no. 4, pp. 210–224, 2005.

[46] D. Paraskevis, E. Magiorkinis, G. Magiorkinis et al., "Increasing prevalence of HIV-1 subtype a in Greece: estimating epidemic history and origin," *Journal of Infectious Diseases*, vol. 196, no. 8, pp. 1167–1176, 2007.

[47] D. Paraskevis, O. Pybus, G. Magiorkinis et al., "Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach," *Retrovirology*, vol. 6, article 49, 2009.

[48] M. Salemi, T. de Oliveira, M. Ciccozzi, G. Rezza, and M. M. Goodenow, "High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania," *PLoS One*, vol. 3, no. 1, Article ID e1390, 2008.

[49] M. Salemi, M. M. Goodenow, S. Montieri et al., "The HIV type 1 epidemic in Bulgaria involves multiple subtypes and is sustained by continuous viral inflow from West and East European countries," *AIDS Research and Human Retroviruses*, vol. 24, no. 6, pp. 771–779, 2008.

[50] E. C. Holmes, "The phylogeography of human viruses," *Molecular Ecology*, vol. 13, no. 4, pp. 745–756, 2004.

[51] M. Slatkin and W. P. Maddison, "A cladistic measure of gene flow inferred from the phylogenies of alleles," *Genetics*, vol. 123, no. 3, pp. 603–613, 1989.

[52] A. R. Templeton, "Statistical phylogeography: methods of evaluating and minimizing inference errors," *Molecular Ecology*, vol. 13, no. 4, pp. 789–809, 2004.

[53] J. Parker, A. Rambaut, and O. G. Pybus, "Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty," *Infection, Genetics and Evolution*, vol. 8, no. 3, pp. 239–246, 2008.

[54] P. Lemey, A. Rambaut, A. J. Drummond, and M. A. Suchard, "Bayesian phylogeography finds its roots," *PLoS Computational Biology*, vol. 5, no. 9, Article ID e1000520, 2009.

[55] P. Lemey, A. Rambaut, J. J. Welch, and M. A. Suchard, "Phylogeography takes a relaxed random walk in continuous space and time," *Molecular Biology and Evolution*, vol. 27, no. 8, pp. 1877–1885, 2010.

[56] S. Kumar, "Molecular clocks: four decades of evolution," *Nature Reviews Genetics*, vol. 6, no. 8, pp. 654–662, 2005.

[57] A. Rambaut, "Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies," *Bioinformatics*, vol. 16, no. 4, pp. 395–399, 2000.

[58] A. J. Drummond, O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo, "Measurably evolving populations," *Trends in Ecology and Evolution*, vol. 18, no. 9, pp. 481–488, 2003.

[59] A. J. Drummond, S. Y. Ho, M. J. Phillips, and A. Rambaut, "Relaxed phylogenetics and dating with confidence," *PLoS biology*, vol. 4, no. 5, article e88, 2006.

[60] H. Kishino, J. L. Thorne, and W. J. Bruno, "Performance of a divergence time estimation method under a probabilistic model of rate evolution," *Molecular Biology and Evolution*, vol. 18, no. 3, pp. 352–361, 2001.

[61] M. J. Sanderson, "A nonparametric approach to estimating divergence times in the absence of rate constancy," *Molecular Biology and Evolution*, vol. 14, no. 12, pp. 1218–1231, 1997.

[62] J. L. Thorne, H. Kishino, and I. S. Painter, "Estimating the rate of evolution of the rate of molecular evolution," *Molecular Biology and Evolution*, vol. 15, no. 12, pp. 1647–1657, 1998.

[63] A. D. Yoder and Z. Yang, "Estimation of primate speciation dates using local molecular clocks," *Molecular Biology and Evolution*, vol. 17, no. 7, pp. 1081–1090, 2000.

[64] L. Bromham and D. Penny, "The modern molecular clock," *Nature Reviews Genetics*, vol. 4, no. 3, pp. 216–224, 2003.

[65] A. D. Yoder and Z. Yang, "Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context," *Molecular Ecology*, vol. 13, no. 4, pp. 757–773, 2004.

[66] M. J. Sanderson, "Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach," *Molecular Biology and Evolution*, vol. 19, no. 1, pp. 101–109, 2002.

[67] B. Rannala and Z. Yang, "Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference," *Journal of Molecular Evolution*, vol. 43, no. 3, pp. 304–311, 1996.

[68] P. Lemey, A. Rambaut, and O. G. Pybus, "HIV evolutionary dynamics within and among hosts," *AIDS Reviews*, vol. 8, no. 3, pp. 125–140, 2006.

[69] B. Korber, M. Muldoon, J. Theiler et al., "Timing the ancestor of the HIV-1 pandemic strains," *Science*, vol. 288, no. 5472, pp. 1789–1796, 2000.

[70] P. Lemey, O. G. Pybus, W. Bin, N. K. Saksena, M. Salemi, and A. M. Vandamme, "Tracing the origin and history of the HIV-2 epidemic," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 11, pp. 6588–6592, 2003.

[71] M. Worobey, M. Gemmel, D. E. Teuwen et al., "Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960," *Nature*, vol. 455, no. 7213, pp. 661–664, 2008.

[72] J. F. C. Kingman, "On the genealogy of large populations," *Journal of Applied Probability*, vol. 19, pp. 27–43, 1982.

[73] R. C. Griffiths and S. Tavaré, "Sampling theory for neutral alleles in a varying environment," *Philosophical transactions of the Royal Society of London. Series B*, vol. 344, no. 1310, pp. 403–410, 1994.

[74] A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus, "Bayesian coalescent inference of past population dynamics from molecular sequences," *Molecular Biology and Evolution*, vol. 22, no. 5, pp. 1185–1192, 2005.

[75] O. G. Pybus, A. Rambaut, and P. H. Harvey, "An integrated framework for the inference of viral population history from reconstructed genealogies," *Genetics*, vol. 155, no. 3, pp. 1429–1437, 2000.

[76] K. Strimmer and O. G. Pybus, "Exploring the demographic history of DNA sequences using the generalized skyline plot," *Molecular Biology and Evolution*, vol. 18, no. 12, pp. 2298–2305, 2001.

[77] O. G. Pybus, E. C. Holmes, and P. H. Harvey, "The mid-depth method and HIV-1: a practical approach for testing hypotheses of viral epidemic history," *Molecular Biology and Evolution*, vol. 16, no. 7, pp. 953–959, 1999.

[78] A. J. Drummond, G. K. Nicholls, A. G. Rodrigo, and W. Solomon, "Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data," *Genetics*, vol. 161, no. 3, pp. 1307–1320, 2002.

[79] V. N. Minin, E. W. Bloomquist, and M. A. Suchard, "Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics," *Molecular Biology and Evolution*, vol. 25, no. 7, pp. 1459–1471, 2008.

[80] A. J. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis by sampling trees," *BMC Evolutionary Biology*, vol. 7, no. 1, article 214, 2007.

[81] J. Heled and A. J. Drummond, "Bayesian inference of population size history from multiple loci," *BMC Evolutionary Biology*, vol. 8, no. 1, article 289, 2008.

[82] M. T. P. Gilbert, A. Rambaut, G. Wlasiuk, T. J. Spira, A. E. Pitchenik, and M. Worobey, "The emergence of HIV/AIDS in the Americas and beyond," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 47, pp. 18566–18570, 2007.

[83] K. E. Potts, M. L. Kalish, T. Lott et al., "Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. Brazilian Collaborative AIDS Research Group," *AIDS*, vol. 7, pp. 1191–1197, 1993.

[84] E. Leal and F. E. Villanova, "Diversity of HIV-1 subtype B: implications to the origin of BF recombinants," *PLoS One*, vol. 5, no. 7, Article ID e11833, 2010.

[85] D. T. Covas, T. A. Bíscaro, S. Kashima, G. Duarte, and A. A. Machado, "High frequency of the GWG (Pro Trp) envelope variant of HIV-1 in Southeast Brazil," *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, vol. 19, no. 1, pp. 74–79, 1998.

[86] M. G. Morgado, M. L. Guimarães, I. Neves Júnior et al., "Molecular epidemiology of HIV in Brazil: polymorphism of the antigenically distinct HIV-1 B subtype strains. The Hospital Evandro Chagas AIDS Clinical Research Group," *Memórias do Instituto Oswaldo Cruz*, vol. 93, no. 3, pp. 383–386, 1998.

[87] G. Bello, W. A. Eyer-Silva, J. C. Couto-Fernandez et al., "Demographic history of HIV-1 subtypes B and F in Brazil," *Infection, Genetics and Evolution*, vol. 7, no. 2, pp. 263–270, 2007.

[88] M. E. Pinto, C. G. Schrago, A. B. Miranda, and C. A. M. Russo, "A molecular study on the evolution of a subtype B variant frequently found in Brazil," *Genetics and Molecular Research*, vol. 7, no. 4, pp. 1031–1044, 2008.

[89] R. S. Diaz, E. Leal, S. Sanabani et al., "Selective regimes and evolutionary rates of HIV-1 subtype B V3 variants in the Brazilian epidemic," *Virology*, vol. 381, no. 2, pp. 184–193, 2008.

[90] P. C. Aulicino, G. Bello, C. Rocco et al., "Description of the first full-length HIV type 1 subtype F1 strain in Argentina: implications for the origin and dispersion of this subtype in South America," *AIDS Research and Human Retroviruses*, vol. 23, no. 10, pp. 1176–1182, 2007.

[91] M. L. Guimarães, A. C. P. Vicente, K. Otsuki et al., "Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates," *Retrovirology*, vol. 6, article 39, 2009.

[92] D. A. Dilernia, L. R. Jones, M. A. Pando et al., "Analysis of HIV type 1 BF recombinant sequences from south america dates the origin of CRF12-BF to a recombination event in the 1970s," *AIDS Research and Human Retroviruses*, vol. 27, no. 5, pp. 569–578, 2011.

[93] S. R. Mehta, J. O. Wertheim, W. Delport et al., "Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania," *Infection, Genetics and Evolution*, vol. 11, no. 5, pp. 975–979, 2011.

[94] M. Gomez Carrillo, M. Avila, J. Hierholzer et al., "Mother-to-child HIV type 1 transmission in Argentina: BF recombinants have predominated in infected children since the mid-1980s," *AIDS Research and Human Retroviruses*, vol. 18, no. 7, pp. 477–483, 2002.

[95] P. C. Aulicino, E. C. Holmes, C. Rocco, A. Mangano, and L. Sen, "Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina," *Journal of Virology*, vol. 81, no. 1, pp. 427–429, 2007.

[96] G. Bello, P. C. Aulicino, D. Ruchansky et al., "Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay," *Retrovirology*, vol. 7, article 22, 2010.

[97] N. Ristic, J. Zukurov, W. Alkmim, R. S. Diaz, L. M. Janini, and M. P.S. Chin, "Analysis of the origin and evolutionary history of HIV-1 CRF28_BF and CRF29_BF reveals a decreasing prevalence in the AIDS epidemic of Brazil," *PLoS One*, vol. 6, no. 3, Article ID e17485, 2011.

[98] G. Bello, C. P. B. Passaes, M. L. Guimarães et al., "Origin and evolutionary history of HIV-1 subtype C in Brazil," *AIDS*, vol. 22, no. 15, pp. 1993–2000, 2008.

[99] R. Fontella, M. A. Soares, and C. G. Schrago, "On the origin of HIV-1 subtype C in South America," *AIDS*, vol. 22, no. 15, pp. 2001–2011, 2008.

[100] T. de Oliveira, D. Pillay, and R. J. Gifford, "The HIV-1 subtype C epidemic in South America is linked to the United Kingdom," *PLoS One*, vol. 5, no. 2, Article ID e9311, 2010.

[101] N. M.C. Véras, R. R. Gray, L. F.M. Brígido, R. Rodrigues, and M. Salemi, "High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America," *Journal of General Virology*, vol. 92, no. 7, pp. 1698–1709, 2011.

[102] L. R. Jones, D. A. Dilernia, J. M. Manrique, F. Moretti, H. Salomón, and M. Gomez-Carrillo, "In-depth analysis of the origins of HIV type 1 subtype C in South America," *AIDS Research and Human Retroviruses*, vol. 25, no. 10, pp. 951–959, 2009.

[103] L. F. D. M. Brigido, "On the origin of South America HIV-1 C epidemic," *AIDS*, vol. 23, no. 4, pp. 543–544, 2009.

[104] R. Fontella, M. A. Soares, and C. G. Schrago, "The origin of South American HIV-1 subtype C: lack of evidence for a Mozambican ancestry," *AIDS*, vol. 23, no. 14, pp. 1926–1928, 2009.

[105] M. Salemi, T. de Oliveira, M. A. Soares et al., "Different epidemic potentials of the HIV-1B and C subtypes," *Journal of Molecular Evolution*, vol. 60, no. 5, pp. 598–605, 2005.

[106] A. F. Santos, C. G. Schrago, A. M. B. Martinez et al., "Epidemiologic and evolutionary trends of HIV-1 CRF31_BC-related strains in southern Brazil," *Journal of Acquired Immune Deficiency Syndromes*, vol. 45, no. 3, pp. 328–333, 2007.

[107] C. P. B. Passaes, G. Bello, R. S. Lorete et al., "Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in Southern Brazil," *Infection, Genetics and Evolution*, vol. 9, no. 4, pp. 474–482, 2009.

[108] B. Korber, J. Theiler, and S. Wolinsky, "Limitations of a molecular clock applied to considerations of the origin of HIV-1," *Science*, vol. 280, no. 5371, pp. 1868–1871, 1998.

[109] T. K. Seo, J. L. Thorne, M. Hasegawa, and H. Kishino, "A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times," *Bioinformatics*, vol. 18, no. 1, pp. 115–123, 2002.

[110] G. Bello, M. L. Guimarães, S. L. Chequer-Fernandez et al., "Increasing genetic distance to HIV-1 subtype B and F1

consensus sequences in the Brazilian epidemic: a challenge for vaccine strategies based on central immunogens?" *Infection, Genetics and Evolution*, vol. 7, no. 5, pp. 594–599, 2007.

[111] G. Bello, M. L. Guimarães, C. P.B. Passaes, S. E.M. Almeida, V. G. Veloso, and M. G. Morgado, "Short communication: evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31-BC epidemics in southern Brazil," *AIDS Research and Human Retroviruses*, vol. 25, no. 11, pp. 1065–1069, 2009.

[112] Brazilian Ministry of Health. AIDS Epidemiological Bulletin [in Portuguese], November 2007.

[113] K. E. Robbins, P. Lemey, O. G. Pybus et al., "U.S. human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains," *Journal of Virology*, vol. 77, no. 11, pp. 6359–6366, 2003.

[114] S. Hué, D. Pillay, J. P. Clewley, and O. G. Pybus, "Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 12, pp. 4425–4429, 2005.

[115] P. R. Walker, O. G. Pybus, A. Rambaut, and E. C. Holmes, "Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth," *Infection, Genetics and Evolution*, vol. 5, no. 3, pp. 199–208, 2005.

[116] G. Zehender, E. Ebranati, A. Lai et al., "Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy," *Journal of Acquired Immune Deficiency Syndromes*, vol. 55, no. 2, pp. 156–160, 2010.

[117] H. F. Njai, Y. Gali, G. Vanham et al., "The predominance of Human Immunodeficiency Virus type I (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness," *Retrovirology*, vol. 3, article 40, 2006.

[118] R. M. Brindeiro, R. S. Diaz, E. C. Sabino et al., "Brazilian network for HIV drug resistance surveillance (HIV-BResNet): a survey of chronically infected individuals," *AIDS*, vol. 17, no. 7, pp. 1063–1069, 2003.

[119] L. F. D. M. Brígido, H. M. Franco, R. M. Custódio et al., "Molecular characteristics of HIV type 1 circulating in São Paulo, Brazil," *AIDS Research and Human Retroviruses*, vol. 21, no. 7, pp. 673–682, 2005.

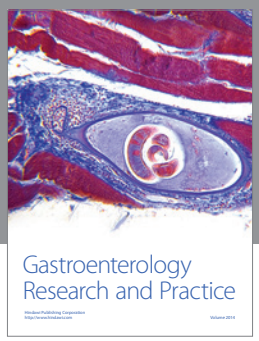
[120] M. L. Guimarães, J. C. Couto-Fernandez, W. D. A. Eyer-Silva, S. L. M. Teixeira, S. L. Chequer-Fernandez, and M. G. Morgado, "Analysis of HIV-1 BF pr/rt recombinant strains from Rio de Janeiro/Brazil reveals multiple unrelated mosaic structures," *Infection, Genetics and Evolution*, vol. 10, no. 7, pp. 1094–1100, 2010.

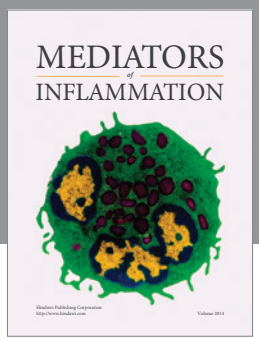
[121] P. C. Aulicino, G. Bello, M. L. Guimaraes et al., "Longitudinal analysis of HIV-1 BF1 recombinant strains in vertically infected children from Argentina reveals a decrease in CRF12_BF pol gene mosaic patterns and high diversity of BF unique recombinant forms," *Infection, Genetics and Evolution*, vol. 11, no. 2, pp. 349–357, 2011.

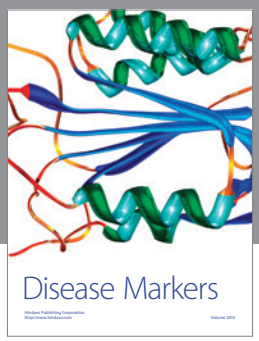
[122] D. J. De Sa-Filho, M. D. S. Soares, V. Candido et al., "HIV type 1 pol gene diversity and antiretroviral drug resistance mutations in Santos, Brazil," *AIDS Research and Human Retroviruses*, vol. 24, no. 3, pp. 347–353, 2008.

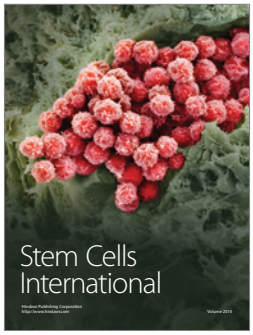
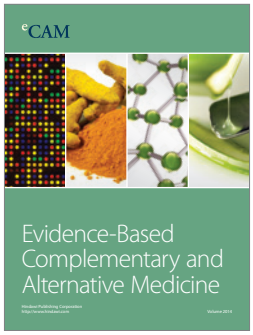
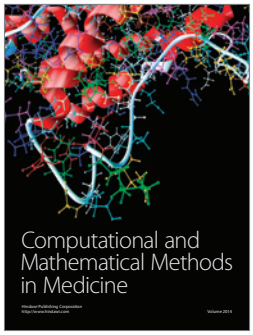
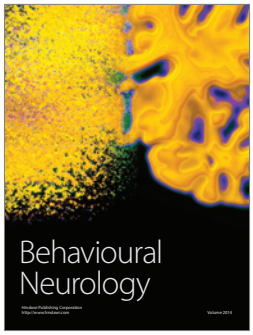
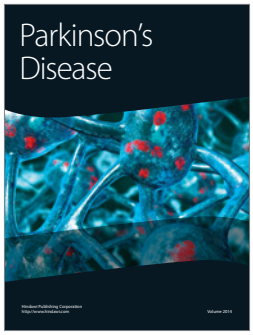
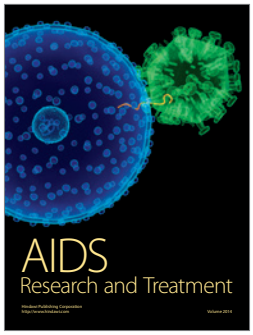
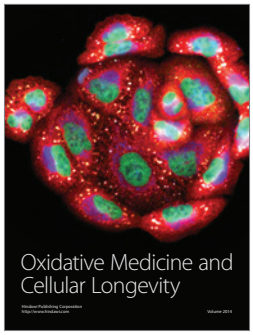
[123] L. F. M. Brígido, C. C. Nunes, C. M. Oliveira et al., "HIV type 1 subtype C and CB pol recombinants prevail at the cities with the highest AIDS prevalence rate in Brazil," *AIDS Research and Human Retroviruses*, vol. 23, no. 12, pp. 1579–1585, 2007.

[124] R. Rodrigues, S. Manenti, P. R. T. Romao et al., "Young pregnant women living with HIV/AIDS in criciuma, Southern Brazil, are infected almost exclusively with HIV type 1 clade C," *AIDS Research and Human Retroviruses*, vol. 26, no. 3, pp. 351–357, 2010.







Hindawi

Submit your manuscripts at
http://www.hindawi.com