

Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region

Rodrigo Cunha Oliveira,¹ Tiago Gräf,² Filipe Ferreira de Almeida Rego,³
Gabriela Porto Santos Almeida Silva,¹ Marta Giovanetti,^{4,5} and Joana Paixão Monteiro Cunha^{1,i}

Abstract

The subtype C accounts for >50% of HIV type 1 (HIV-1) infections worldwide and it is currently the predominant viral form in South Brazil. Subtype C has been reported in all Brazilian regions; however, the phylogenetic relationship among strains circulating in those regions still remains unclear. This study aimed to investigate the origin and dynamic dispersion of HIV-1 subtype C toward Northeast Brazil. Our phylogenetic analysis suggests that most subtype C strains circulating in Brazil (99%) are descendant from the main lineage whose entrance in the country was previously described in the 1970s. According to the literature, additional introductions of subtype C were reported in the country through the Southeast region and in this study we identified another entry event that occurred most likely through the North region. Furthermore, our analysis suggests that the spread of subtype C to Brazilian Northeastern states occurred through multiple independent introductions of the main lineage that originated in South Brazil between mid-1980s and late 1990s. Despite the observation of eventual new HIV-1 subtype C introductions, our results highlight the predominance of a single lineage of this subtype in Brazil and the importance of South region in its dissemination throughout the country.

Keywords: HIV-1, subtype C, Brazil, Northeast

Introduction

INFECTION WITH HIV type 1 (HIV-1) usually results in a progressive destruction of the immune system, establishing a state of immunodepression that allows the appearance of opportunistic infections.¹ Although the advent of anti-retroviral therapy has increased the expectancy and quality of life of HIV-1-infected individuals, there is still no cure or vaccine capable of preventing infection.² The intense genetic variability and capacity to establish reservoirs *in vivo* are the main obstacles in the fight against HIV/AIDS.³ The HIV-1 genetic diversity is represented by the existence of two viruses (HIV-1 and HIV-2), groups, subtypes, and recombinant forms, which have different dispersion patterns around the world.⁴

The HIV-1 epidemic in Brazil is predominantly characterized by the cocirculation of B, C, and F1 subtypes and the BF1 and BC recombinant forms.⁵ Owing to the continental

extent of the country, the Brazilian geographic regions present unique patterns of subtype distribution. Subtypes B and F1 predominate in the North region, whereas B, F1, and BF recombinants are the most frequent forms in the Northeast.^{6,7} In the Southeast and Central West regions, B, C, F1, and BF1 cocirculate,^{8,9} whereas in the South region, HIV-1 infections are caused mainly by subtypes B, C, and BC recombinants.⁹

The subtype C is the most prevalent form of HIV-1 worldwide and it was first recognized in South Brazil by early 1990s, in the Rio Grande do Sul state. Initial studies in that region showed that 22% of the HIV infections were caused by subtype C between 1994 and 1997,¹⁰ whereas most recent reports described a prevalence of 44%–66%,¹¹ pointing to an increase of subtype C prevalence in South Brazil over the last two decades, as previously reviewed.¹² Furthermore, the subtype C prevalence has also been reported in Southeast (16%),¹³ Central West (11.5%),⁸ Northeast (5%),¹⁴ and

¹Núcleo de Bioinformática, Departamento de Bioquímica e Biofísica, Universidade Federal da Bahia, Salvador, Brazil.

²Instituto Gonçalo Moniz, Fundação Oswaldo Cruz—FIOCRUZ, Salvador, Brazil.

³Universidade Católica do Salvador, Salvador, Brazil.

⁴Instituto Oswaldo Cruz, Fundação Oswaldo Cruz—FIOCRUZ, Rio de Janeiro, Brazil.

⁵Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

ⁱORCID ID (<https://orcid.org/0000-0001-9074-457X>).

North (2.9%)¹⁵ Brazilian regions. Previous studies have reconstructed the history and the dynamics of the subtype C epidemic in Brazil. A close phylogenetic relationship has been demonstrated between strains circulating in Brazil and sequences from countries located in the eastern region of the African continent, such as Burundi and Kenya.^{16–18}

The Northeast Brazil consists of nine states and is the third region with the highest number of HIV-1 infections in the country.¹⁹ The increase of HIV-1 cases in this region seems to be related to factors such as low economic and educational levels, sex tourism, and prostitution, which might have facilitated the introduction and the spread of new viral forms. Despite previous studies already reporting the circulation of subtype C in Northeast Brazil,^{7,14,20,21} the origin of these viruses remains unknown.

This study uses nucleotide sequences from all five Brazilian regions and aims to reconstruct the virus migration movement to the Northeast region, which has the largest number of federative units and coastal extension, attracting tourists from the other four regions. The generated data highlight the importance of molecular surveillance studies for the development of measures to control HIV-1 spread and for providing important information for the development of prophylactic agents.

Material and Methods

Subtype C Brazilian dataset

To characterize the HIV-1 subtype C epidemic in Northeastern Brazil, a search for sequences from this region was carried out in the Los Alamos Database (www.hiv.lanl.gov). Sequences were selected based on availability of information regarding year and location of sampling. Sequences with low quality and high frequency of degenerate bases were excluded.

Twenty-two sequences from the Northeast region were identified covering the genomic fragment of *pol* gene corre-

sponding to positions 2301–3200 relative to HXB2 reference strain. To compose the Brazilian dataset, a total of 948 subtype C sequences covering the genomic fragment aforementioned were initially collected. Subsequently, sequences were excluded according to the following criteria: absence of information regarding the year and location of sampling, duplicated entries, clones from the same sample, and sequences with high rate of gap, stop codons, and degenerate bases. All subtype C sequences were analyzed through the jpHMM tool²² to exclude possible recombinant samples. The final Brazilian dataset had 460 subtype C sequences from five geographic regions (South $n=336$, Southeast $n=63$, Central West $n=32$, North $n=7$, and Northeast $n=22$) isolated between 1998 and 2017 (Fig. 1).

Subtype C reference dataset

To identify the different HIV-1 subtype C strains circulating in Brazil, a consensus sequence of Brazilian dataset ($n=460$) was created in the GENEIOUS software²³ and submitted to HIV BLAST online tool against all subtype C sequences (https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). A total of 200 hit sequences were collected and repeated sequences from Brazil were excluded. After manual inspection, sequences were excluded according to the criteria described previously. The final subtype C reference dataset contained 110 samples from the following countries: Botswana, India, Kenya, Malawi, Senegal, Somalia, South Africa, Tanzania, Uganda, United States, and Zambia.

Sequence alignment and phylogenetic analyses

Alignment was performed using MAFFT online program²⁴ (<http://mafft.cbrc.jp>) under the command: `mafft —thread 8 —threadtb 5 —threadit 0 —reorder —auto input > output` and manually edited using BioEdit software.²⁵ The dataset

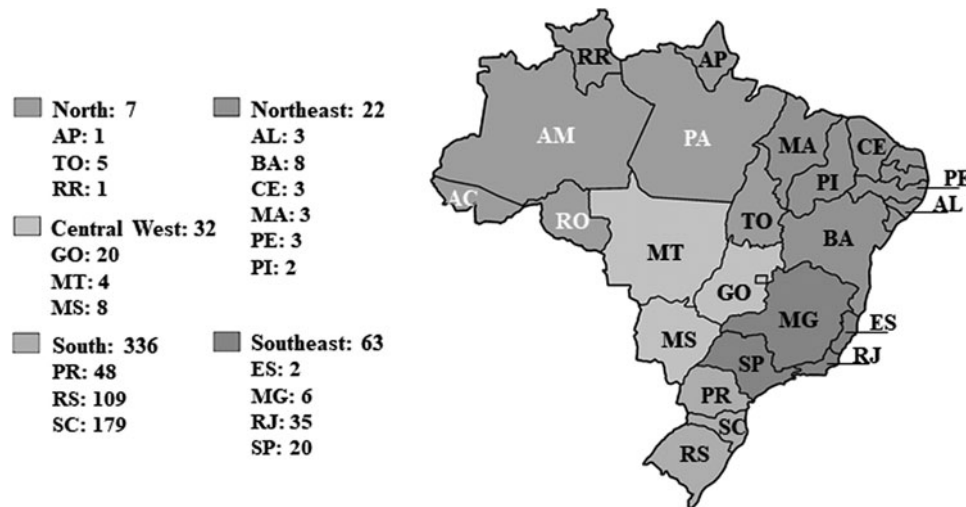


FIG. 1. Distribution map of analyzed sequences by Brazilian regions and states ($n=460$). The geographic regions are identified by different shades: North, Northeast, Central West, Southeast and South. The states are indicated by two letters code: AL (Alagoas), AP (Amapá), BA (Bahia), CE (Ceará), ES (Espírito Santo), GO (Goiás), MG (Minas Gerais), MS (Mato Grosso do Sul), MT (Mato Grosso), PE (Pernambuco), PI (Piauí), PR (Paraná), RJ (Rio de Janeiro), RS (Rio Grande do Sul), SC (Santa Catarina), SP (São Paulo), and TO (Tocantins). Unrepresented states are labeled in *white*: AC (Acre), AM (Amazonas), PA (Pará), and RO (Rondônia). The number of HIV-1 subtype C *pol* sequences from each location included in the present study are indicated. HIV-1, HIV type 1.

was assessed for presence of phylogenetic signal by applying the likelihood mapping analysis implemented in the TreePuzzle Program.²⁶

Maximum likelihood (ML) phylogenies were reconstructed using IQ-TREE 1.6.8 webserver.²⁷ The reconstruction was performed under the GTR+I+G nucleotide substitution model that was inferred in Modelfinder.²⁸ Subtree pruning–regrafting branch-swapping algorithm was used to perform Heuristic tree search. The reliability of each cluster was evaluated by analyzing 1,000 bootstrap (BS) replicates and with the approximate likelihood ratio test (aLRT)²⁹ based on the Shimodaira–Hasegawa-like procedure. The ML trees were visualized by FigTree program version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Analysis of structural motifs and amino acids

For the purpose of investigating and comparing amino acids patterns and signatures, the sequences representing different subtype C introductions were translated in GENEDOC³⁰ in association with the biological information package of the Prosite Database (<https://prosite.expasy.org>). A consensus sequence of the monophyletic group of Brazilian sequences (CONS_C_BR) was created in the GENEIOUS software²³ and the sequences were then compared with the world consensus subtype C obtained from Los Alamos database.

Phylogenetic and phylogeographic dispersion of HIV-1 subtype C toward Northeast region

To investigate the temporal signal of the dataset, we regressed root-to-tip genetic distances from the ML trees against sample collection dates using TempEst v 1.5.3.³¹ The Bayesian analyses were conducted in BEAST 1.10.4 package.^{32–34} The ML phylogeny was used as a starting tree for Bayesian time-scaled phylogenetic analysis using BEAST 1.10.4 in CIPRES SCIENCE GATEWAY.^{32,33,35} The tree reconstruction was inferred with the GTR + Γ 4 nucleotide substitution model, an uncorrelated relaxed molecular clock, and the demographic model of logistic growth, according to previous findings investigating the fittest models for HIV-1 subtype C in Brazil.³⁶

The percentage of fully resolved probability mapping quartets totals 85.8% and because of the dataset small temporal signal ($r=0.38$), we applied an informative normal prior (mean = 41, SD = 5.1) on the root height based on previous study that estimated the introduction HIV-1 subtype C in Brazil to 1976 (1966–1983, 95% high posterior density).¹⁸ Four runs of Markov chain Monte Carlo (MCMC) with 100 million states each were computed, sampling every 10 million steps. The log and trees files were combined in Log-Combiner discarding 40% as burn-in and the convergence of MCMC chains was checked using Tracer v.1.7.1.³⁷ The value of effective sample size of proper mixing was 282.2 (>200) and the maximum clade credibility (MCC) trees were summarized from the MCMC samples using TreeAnnotator by CIPRES.³⁵

To estimate the transmission dynamics of subtype C toward the Northeast region of Brazil, phylogeographic analysis was performed on the set of empirical trees obtained by the Bayesian phylogenetic analysis. A discrete diffusion model with Bayesian stochastic search variable selection as

implemented in BEAST was applied to reconstruct the ancestral location states of the trees. Tip locations were defined as the Brazilian geographic region from which each sequence originated: North, Northeast, Central West, and Southeast regions. Owing to the large sample size, sequences from the South region were further classified according to the state of origin: Paraná (PR), Rio Grande do Sul (RS), and Santa Catarina (SC). The migratory events were summarized and visualized using the SPREAD3 tool.³⁸

Results

Phylogenetic inference of the HIV-1 subtype C epidemic in Brazil and introduction in Northeast region

To investigate the evolutionary history of subtype C in Brazil, a preliminary ML tree was reconstructed using 460 Brazilian subtype C sequences plus 110 subtype C sequences worldwide (Fig. 2). Full details of the sequences analyzed in this study are provided in the Supplementary Material S1. No recombination events were found among the sequences. The phylogenetic reconstruction showed that 455 (98.9%) Brazilian subtype C sequences fall within a single well-supported major clade (BS = 88; aLRT = 99) comprising isolates from all five Brazilian regions sampled between 1998 and 2015. Moreover, five Brazilian viral sequences grouped separately from the main monophyletic clade, clustering with subtype C sequences from other countries (Kenia, Senegal, South Africa, and Uganda), were supported by both BS (70–89 interval) and aLRT (47–99 interval) values. Of note, all sequences from Northeast region branched within the larger Brazilian subtype C clade (Fig. 2).

The presence of characteristic amino acid substitutions was investigated among the sequences representing the six putative subtype C introduction in Brazil (Supplementary Material S2). In the fragment of *pol* analyzed here, 14 motifs were conserved in all sequences. These predicted regions correspond to cyclic adenosine monophosphate or guanosine monophosphate kinase protein phosphorylation sites, Casein kinase II and tyrosine kinase phosphorylation sites, active sites of protease, *N*-myristoylation and amidation sites.

Considering only the substitutions with frequency $\geq 75\%$, the consensus sequence (CONS_C_BR) of the major Brazilian monophyletic cluster ($n=455$) differed at five positions from the world consensus C [protease: I75L, N93K, K97N; reverse transcriptase (RT): E92D and G174D], five of which represent exclusive signatures of this group. The other five samples representing additional entries of HIV-1 subtype C into Brazil (KF255847, KF255848, KF255855, KF255861, and KX443060) have at least five unique amino acid signatures each, thereby showing different signature patterns (Supplementary Material S2).

Inference of HIV-1 subtype C spread toward Northeast region

To further investigate the introduction events and the dynamic dispersion of subtype C toward the Northeast region, phylogeographic analyses were conducted. For these analyses, the dataset composed of 455 sequences branched inside the Brazilian monophyletic clade of subtype C.

ML analysis showed that the 22 subtype C sequences from Northeast region appeared to be distributed among 19

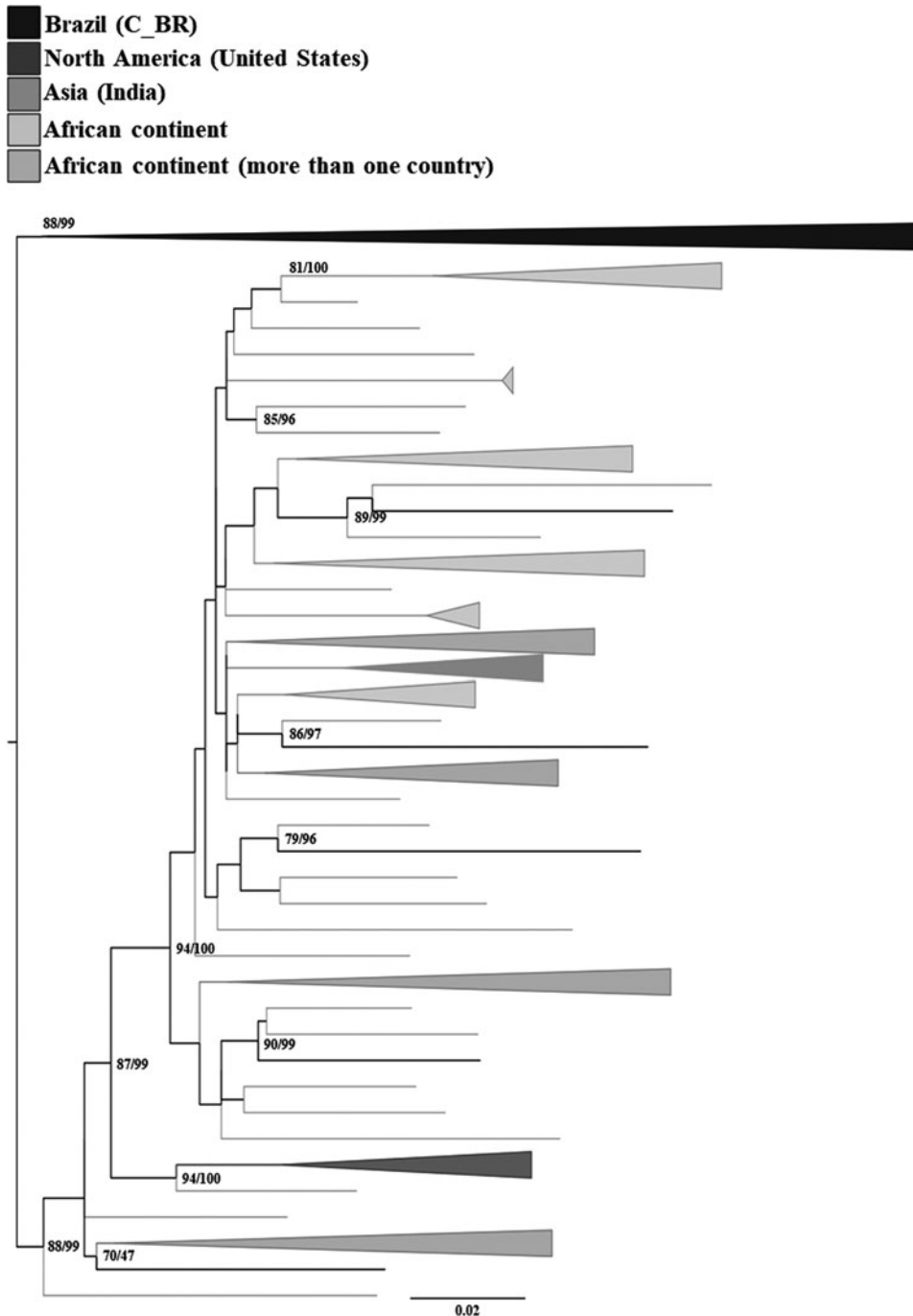


FIG. 2. Maximum likelihood analysis of HIV-1 subtype C pol sequences (876bp) from Brazil ($n=460$) and worldwide ($n=110$). The color of the branches represents the geographic origin of the subtype C sequences. A major monophyletic group formed by subtype C sequences from Brazil is shown as a collapsed cluster (455) while five Brazilian samples grouped outside this cluster. The clusters formed by sequences from Africa, Asia (India), and the North America are collapsed and presented in different colors. The tree was built under the evolutionary model GTR+I+G and visualized in the software Figtree v1.4.4. The statistical support is indicated only at key nodes as bootstrap/aLRT values. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating 0.02 nucleotide substitutions per site. aLRT, approximate likelihood ratio test.

independent lineages that were intermixed between the sequences from other Brazilian regions (Supplementary Material S3). All 19 clusters involving Northeastern sequences were well supported with both BS and aLRT values, except one clade (BS=47/aLRT=46), which grouped a sequence from Alagoas state (Northeast) with a sequence from Mato Grosso do Sul (Central West). Most subtype C lineages from Northeast Brazil composed of only one sequence from this location and only three clades grouped two sequences from this region. Of note, most clades containing sequences from the Northeast also had sequences from two, three, or four other geographic regions (Supplementary Material S3).

To estimate the time of subtype C introductions in Northeast region, Bayesian analyzes were conducted (Fig. 3). The estimated evolutionary rate was 2.068×10^{-3} substitutions/sites/year (95% Bayesian Credibility Interval $1.0-2.6 \times 10^{-3}$), consistent with other estimates for this subtype.^{17,18} In this reconstruction, sequences from Northeast were distributed in the tree along with sequences from South, Southeast, North, and Central West regions (Fig. 3). The topologies in this MCC tree were remarkably similar to those in the ML analysis (Supplementary Material S3). The clades formed by sequences from the Northeast presenting PP >0.8 had the time to the most recent common ancestor (tMRCAs)

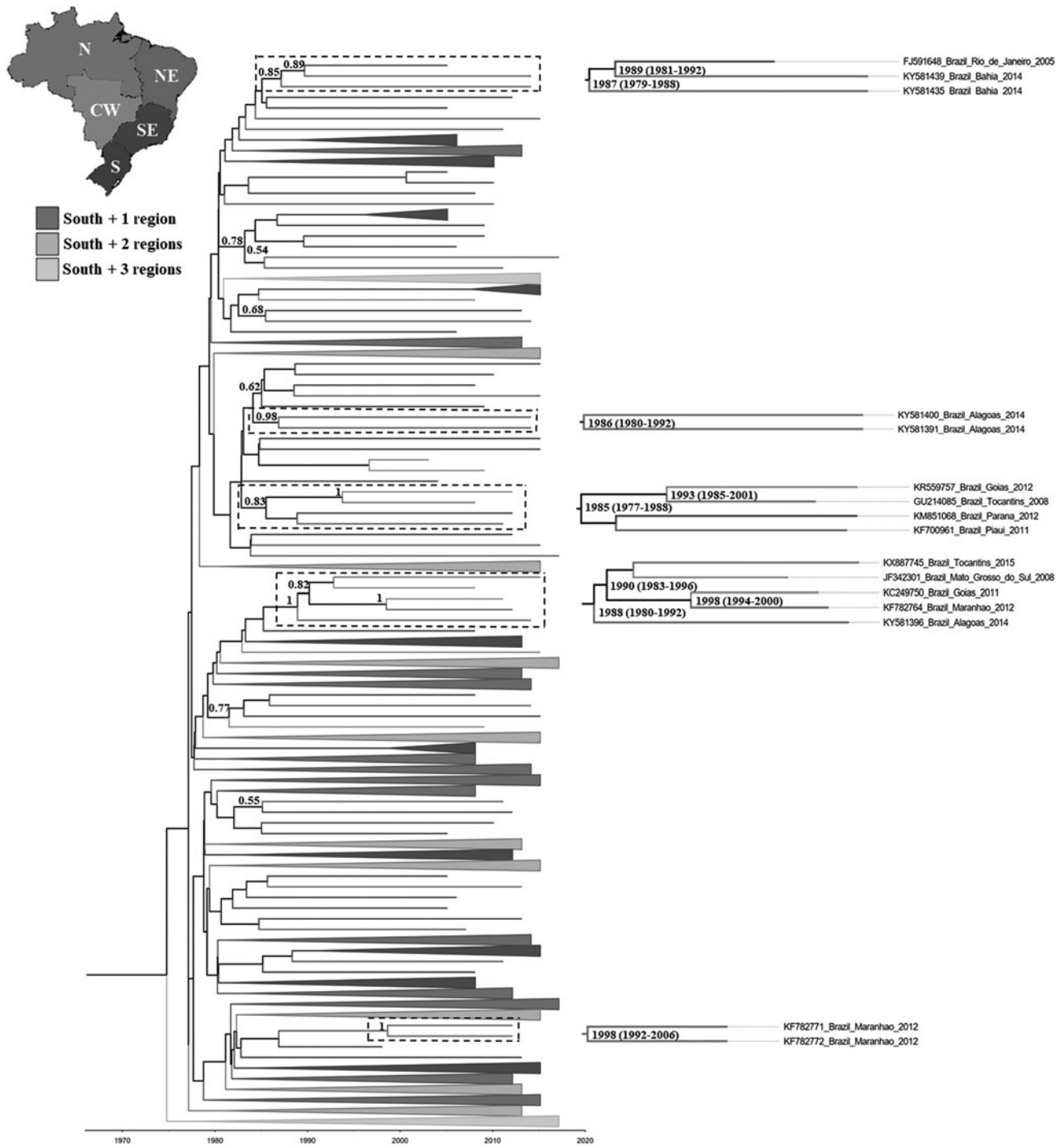


FIG. 3. Time scale Bayesian maximum clade credibility tree showing phylogenetic relationships among subtype C viruses circulating in Brazil ($n=455$). North (6), Northeast (22), Central West (32), Southeast (59), and South (336). Branch colors represent the geographic region from where the subtype C strain originated, according to the legend given in the figure. The collapsed clusters represent groups formed by sequences from the South region plus one, two or three other regions. The posterior probability (PP) is indicated only at key nodes and the dotted boxes highlight those with a high (>0.80) PP support. The median age (with 95% high posterior density interval in parentheses) for subtype C migration events into Northeast region are shown. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating years.

TABLE 1. SUBTYPE C DATA IN THE NORTHEAST REGION BASED ON PHYLODYNAMIC TREE

Posterior probability	Year	Confidence interval
0.83	1985	1977–1988
0.85	1987	1979–1988
1	1988	1980–1992
0.89	1989	1981–1992
0.82	1990	1983–1996
1	1993	1985–2001
1	1998	1994–2000

estimated (Fig. 3 and Table 1). The oldest node (PP=0.83) showing evidence of subtype C transmission to the Northeast region (Piauí state) contained sequences from three other regions: South (Paraná state), Central West (Goiás state), and North (Tocantins state) with the tMRCA=1985 (1977–1988). A second branch (PP=0.85) grouped two sequences from Bahia state (Northeast region) and one from Rio de Janeiro state (Southeast region) [tMRCA=1987 (1979–1988)]. The third branch (PP=1) was formed by sequences from Northeast (Alagoas and Maranhão states), Central West (Goiás and Mato Grosso do Sul states), and North (Tocantins) with the tMRCA estimated in 1988 (1980–1992).

Only two nodes contained two sequences from the same state of Northeast region. The first node contained two sequences from the state of Alagoas, both isolated in 2014 [tMRCA=1986 (1980–1992); PP=0.97] and the second grouped two samples from Maranhão state collected in 2012 [tMRCA=1998 (1992–2006); PP=1] (Fig. 3). The result of phylogeography under SPREAD3 revealed an apparent migratory flow of HIV-1 subtype C mainly between the states of the southern region (Paraná, Santa Catarina, and Rio Grande

do Sul). Subsequently, the dissemination occurred toward the states in the north of the country, reaching the Central West, Southeast, North, and Northeast regions (Fig. 4 and Supplementary Materials S4 and S5). Considering the value of PP>0.80 as a cutoff in the phylogeographic tree, the node information points to the South region as the probable source of the subtype C strains found in the Northeast region.

Discussion

Among 10 different subtypes and 102 circulating recombinant forms of HIV-1, C subtype alone accounts for nearly one half of worldwide infections.³⁹ The high prevalence rates of this subtype are detected in countries from Asia and Africa, mainly in the East and South of the African continent.³⁹ In some countries, subtype C became the prevalent HIV-1 form after its introduction, surpassing another previously dominant genotype, such as in South Africa^{40,41} and India.^{42,43} In South Brazil, subtype C overcame, over the last two decades, the epidemic initially characterized by subtype B predominance and, in addition, it spread to other regions, being reported especially in the neighboring regions, Southeast and Central West.^{11,18}

This study investigated the phylogenetic relationships among subtype C strains isolated from all Brazilian geographic regions, representing 19 of the 26 state units in the country. Our analysis showed that majority (98.9%) of the subtype C viruses circulating in Brazil are descendants of a main lineage that, according to previous studies, has been introduced in the country through the South region between 1976 and 1983.^{17,18}

Of the 460 sequences, five Brazilian viral isolates grouped separately from the main major clade, providing evidence that multiple subtype C introductions have occurred in Brazil.

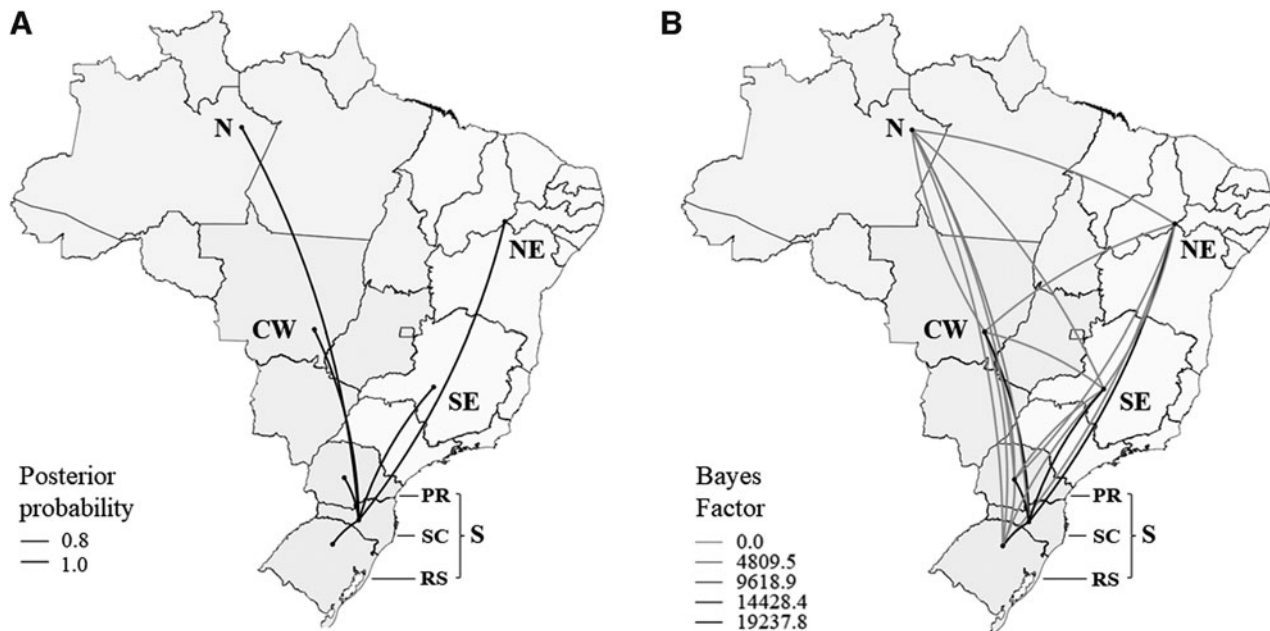


FIG. 4. Viral dispersion dynamics of HIV-1 C in Brazil. Phylogeographic analysis showing subtype C transmission route from South and it to all other Brazilian geographic regions. Viral strains from Northeast originated from South. The lines represent the supported transitions with a PP=1.0 (A) and Bayes factor values (B). The geographic regions are represented by letters: North (N), Northeast (NE), Central West (CW), Southeast (SE), and South (S). The Southern states are indicated (PR—Paraná, RS—Rio Grande do Sul and SC—Santa Catarina). Analysis was performed using SPREAD3.

Apart from these, four lineages have been previously identified, indicating that additional entry events of subtype C have occurred through Southeast region of Brazil.¹⁸ Here, we bring evidence of an additional introduction of subtype C in Brazil represented by a strain isolated in Roraima⁴⁴ (North region), which grouped with a sequence from South Africa (BS = 79; aLRT = 96).

The comparative molecular characterization between the consensus sequence of Brazilian main subtype C clade (CONS_BR_C; $n=455$) and the five sequences outside this group (KF255847, KF255848, KF255855, KF255861, and KX443060) revealed structural motifs associated with phosphorylation sites on *pol* gene. Previous reports have demonstrated that HIV-1 RT is a substrate for *in vitro* phosphorylation by several protein kinases, although its influence on RT activity is not completely understood.^{45,46} According to previous studies, the Brazilian subtype C indeed reveals specific amino acid patterns.^{47,48}

Within the protease and the RT regions of the *pol* gene, the Brazilian subtype C consensus sequence presents four substitutions that were not shared with the consensus sequences from Botswana, India, South Africa, Tanzania, and Zambia.⁴⁷ In addition, 12 specific positions in *gag* and *env* genes were also not found in other consensus sequences from Botswana, Ethiopia, India, South Africa, and Tanzania.⁴⁸ Our results demonstrate that Brazilian consensus (CONS_C_BR) and the five sequences representing additional entries of HIV-1 subtype C in Brazil have five unique amino acid signatures each. Therefore, they exhibit different signature patterns among them. These substitutions in amino acid signatures found in CONS_C_BR and in each sample representing additional subtype C introductions in Brazil reinforces the independent origin of these strains (Supplementary Material S2).

Despite its current predominance in the southern region of Brazil, subtype C has been reported in Northeast with a 5% prevalence.¹⁴ However, the origin and epidemiological relationships of the subtype C variants circulating in that region are unknown. In this study, we collected all the subtype C sequences from the Northeast available in public databases to assess its dynamic dispersion.

The ML reconstructions indicated that all subtype C viruses from Brazilian Northeast region resulted from the dissemination of the main subtype C lineage (C_BR) (Fig. 2), which was introduced multiple times in this part of Brazil (Supplementary Material S3). Although most samples from Northeast are isolated from each other in the phylogenetic trees, they grouped with sequences from the other four geographic regions (South, Southeast, Central West, and North) (Fig. 3 and Supplementary Material S3), pointing to a high level of genetic mixing of Brazilian subtype C viruses from different geographic locations.

Previous Bayesian analyses indicated the contribution of the South region to the subtype C epidemic toward the northernmost states of Brazil.^{18,49} These studies indicated the South region as the main point of initial dispersion of the subtype C in Brazil, especially to Central West region,^{17,18,49} which presents 7.8% of HIV-1 subtype C prevalence.⁹ The factors that would have contributed to the spread of the virus were human mobility owing to geographical proximity and the job opportunities in the fields of agriculture and livestock.^{50,51} In fact, it has been shown that most of the subtype C strains found in Central West region present close phylogenetic

relationships with those isolated in the South.^{8,50} Other studies reported that the Southeast region may also represent an additional strategic point for subtype C dispersion to other regions.^{8,51} Delatorre and colleagues¹⁸ have proposed that the Southeast could represent a secondary point for subtype C dissemination to Central West and back to South.

These observations together with the phylogenetic relations shown in the present analysis (Fig. 3) indicate that Southeast and Central West regions may also constitute HIV-1 subtype C transmission routes toward Northeast, in addition to the main route represented by the South region. Of note, there is one clade composed of sequences from Northeast (Piauí), South (Paraná), North (Tocantins), and Central West (Goiás). This cluster had tMRCA estimated around 1985 (PP = 0.83 95% BCI, 1977–1988) and may represent the oldest introduction of subtype C in Northeast. In combination, these data provide additional evidence that subtype C has expanded from South to northern regions of the country, as previously suggested.⁴⁹ The two nodes containing sequences from Northeast sequences may represent intraregional transmission networks within the same region. In both cases the estimated year is dated after the oldest entry event in this region.

The phylogeographic analysis using SPREAD3 revealed that in the beginning of the 1980s, subtype C was mainly concentrated in the South region, within an intraregional transmission trajectory, whereas an apparent migratory flow of the virus from South to Southeast and Central west regions occurred around the mid-1980s (Fig. 4 A, B and Supplementary Materials S4 and S5). Similarly, Delatorre and collaborators¹⁸ have suggested that viral strains from southern states (Santa Catarina and Paraná) were transmitted to Southeast and Central West regions between 1983 and 1988. Another report indicated that HIV-1 subtype C had already reached the Central West region in 1981 (from Rio Grande do Sul), the Southeast region (São Paulo) during the early 1980s, and the North region (Amazonas and Tocantins states) between late 1980s and early 1990s.⁴⁹

In this sense, there was a gap with respect to the origin and time of subtype C introduction in the Northeast region. The present phylogeographic analysis (Supplementary Materials S4 and S5) suggests that all viral sequences involved in the dispersion of subtype C to Northeast region most likely originated from South region between mid-1980s and late 1990s and that Santa Catarina was the main hub of subtype C dissemination to other locations, corroborating previous studies.^{17,18} However, in the view that Santa Catarina state presents the highest number of sequences in our dataset (Fig. 1), we cannot rule out the impact of sampling heterogeneity in the ancestral states reconstruction of our phylogeographic analysis, as shown in the study by Graf and colleagues.⁴⁹ Therefore, future studies with a more balanced sample distribution should be performed to investigate the HIV-1 dissemination patterns in a finer scale (i.e., states or cities).

Conclusion

The Northeast is the Brazilian region with the highest number of federative units ($n=9$) and the largest coastal extension in the country. According to the Brazilian Ministry of Tourism, the Northeast is the preferred tourist destination for most Brazilians,⁵² attracting millions of visitors every year, 20% of which

are from South and Southeast regions.⁵³ Most subtype C strains circulating in Brazil are descendant of a main lineage; however, this study confirmed the occurrence of additional independent introductions represented by punctual cases. One of these lineages represents evidence of an independent entry of subtype C in Brazil throughout the North region.

The HIV-1 subtype C epidemic in Northeast region resulted from multiple independent introductions of a main Brazilian subtype C lineage that originated in South region between mid-1980s and late 1990s. These Northeastern strains clustered together with viruses isolated from all other four geographic regions, pointing to a high level of genetic mixing of Brazilian subtype C viruses. Taken together, these results provide additional evidence that subtype C has been expanding from the south to the north of the Brazilian territory. These data contribute to the continuous monitoring of the molecular epidemiology of HIV-1, which is of great importance for the surveillance and control of HIV/AIDS epidemic.

Authors' Contributions

R.C.O.: conceptualization, investigation, writing—original draft preparation; T.G.: methodology, formal analysis, writing—review and editing; F.F.A.R.: methodology, validation, writing—review and editing; G.P.S.A.S.: investigation; M.G.: conceptualization, methodology, validation, writing—review and editing; J.P.M.C.: conceptualization, validation, investigation, writing—original draft preparation, supervision.

Acknowledgment

R.C.O. thanks Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for the masters scholarship. M.G. thanks Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). The authors thanks Allan Botura Brennecke Leite for the English proofreading.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

R.C.O. and M.G., were supported by scholarships. This research was not supported by any funding.

Supplementary Material

Supplementary Material S1
Supplementary Material S2
Supplementary Material S3
Supplementary Material S4
Supplementary Material S5

References

- Ioachim HL: Immunopathogenesis of human immunodeficiency virus infection. *Cancer Res* 1990;50:5612–5617.
- Verma R, Khana P, Chawla S, Dhankar M: HIV vaccine: Can it be developed in the 21st century. *Hum Vaccin Immunother* 2015;12:222–224.
- Santoro MM, Perno CF: HIV-1 genetic variability and clinical implications. *ISRN Microbiol* 2013;2013:481314.
- Robertson DL, Anderson JP, Bradac JA: HIV-1 nomenclature proposal. *Science* 2000;288:55–56.
- Silva LA, Divino F, Rêgo MO, *et al.*: HIV-1 Genetic diversity and transmitted drug resistance in anti-retroviral treatment-naïve individuals from Amapá state, Northern Brazil. *AIDS Res Hum Retroviruses* 2016;32:373–376.
- Cunha LKH, Kashima S, Amarante MFC, *et al.*: Distribution of human immunodeficiency virus type 1 subtypes in the state of Amazonas, Brazil, and subtype C identification. *Braz J Med Biol Res* 2012;45:104–112.
- Monteiro-Cunha JP, Souza JSM, Diaz RS, *et al.*: HIV-1 Molecular characterization and transmitted drug resistance prevalence among treatment-naïve individuals. *J AIDS HIV Infect* 2017;3:103–111.
- Alcântara KC, Reis MNG, Cardoso LPV, Bello G, Stefani MMAS: Increasing heterosexual transmission of hiv-1 subtype C in inland Central Western Brazil. *J Med Virol* 2013;85:396–404.
- Arruda MB, Boulosa LT, Cardoso CC, *et al.*: Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): A survey of treatment-naïve individuals. *J Int AIDS Soc* 2018;21:e25032.
- Martínez AM, Barbosa EF, Ferreira PC, *et al.*: Molecular epidemiology of HIV-1 in Rio Grande, RS, Brazil. *Rev Soc Bras Med Trop* 2002;35:471–476.
- Graf T, Fritsch HM, Medeiros RM, Junqueira DM, Almeida SEM, Pinto AR: Comprehensive characterization of HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS. *J Virol* 2016;90:8160–8168.
- Graf T, Pinto AR: The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 2013;435:170–178.
- Alves BM, Siqueira JD, Prellwitz IM, *et al.*: Estimating HIV-1 genetic diversity in Brazil through next-generation sequencing. *Front Microbiol* 2019;10:749.
- Delatorre E, Couto-Fernandez JC, Bello G: HIV-1 genetic diversity in northeastern Brazil: High prevalence of non-B subtypes. *AIDS Res Hum Retroviruses* 2017;33:639–647.
- Machado LFA, Costa IB, Folha MN, *et al.*: Lower genetic variability of HIV-1 and antiretroviral drug resistance in pregnant women from the state of Pará, Brazil. *BMC Infect Dis* 2017;17:270.
- Bello G, Passaes CPB, Guimarães ML, *et al.*: Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS* 2008;22:1993–2000.
- Bello G, Zanotto PMA, Iamarino A, *et al.*: Phylogeographic analysis of hiv-1 subtype c dissemination in Southern Brazil. *PLoS One* 2012;7:e35649.
- Delatorre E, Couto-Fernandez JC, Guimarães ML, *et al.*: Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. *PLoS One* 2013;8:e74072.
- Brazil, Ministry of Health. Health Surveillance Secretary. HIV/AIDS Epidemiological Report. Special number. Available at www.aids.gov.br/pt-br/pub/2019/boletim-epidemiologico-de-hivaids-2019 (2019), accessed June 11, 2020.
- Arruda E, Simões L, Sucupira MC, *et al.*: Short communication: Intermediate prevalence of HIV type 1 primary antiretroviral resistance in Ceará state, Northeast Brazil. *AIDS Res Hum Retroviruses* 2011;27:153–156.

21. Monteiro JP, Alcantara LCJ, Oliveira T, *et al.*: Genetic variability of human immunodeficiency virus-1 in Bahia State, Northeast, Brazil: High diversity of HIV genotypes. *J Med Virol* 2009;81:391–399.
22. Schultz AK, Zang M, Bulla I, *et al.*: jpHMM: Improving the reability of recombination prediction in HIV-1. *Nucleic Acids Res* 2009;37:W647–W651.
23. Kearse M, Moir R, Wilson M, *et al.*: Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.
24. Katoh K, Kuma K, Toh H, Miyata T: MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 2005;33:511–518.
25. Hall T: BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999;41:95–98.
26. Schmidt HA, Petzold E, Vingron M, Haeseler AV: Molecular phylogenetics: Parallelized parameter estimation and quartet puzzling. *J. Parallel Distrib Comput* 2003;63:719–727.
27. Trifinopoulos J, Nguyen LT, Haeseler AV, Minh BQ: W-IQ-TREE: A fast-online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 2016;44:W232–W235.
28. Kalyaanamoorthy S, Minh B, Wong T, Haeseler A, Jermini L: ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
29. Anisimova M, Gascuela O: Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 2006;55:539–552.
30. Nicholas KB: GeneDoc: Analysis and Visualization of Genetic Variation. *EMBnet News* 1997;4:1–4.
31. Rambaut A, Lam TT, Fagundes de Carvalho L, Pybus: Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016;2:vew007.
32. Drummond AJ, Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.
33. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;29:1969–1973.
34. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P: Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 2013;30:239–243.
35. Miller MA, Pfeiffer W, Schwartz T: Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop*. IEEE, G.C.E. New Orleans, LA, 2010, pp. 1–8.
36. Mir D, Graf T, Almeida SEM, Pinto AR, Delatorre E, Bello G: Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches. *Sci Rep* 2018;8:8778.
37. Rambaut A, Drummond AJ, Suchard MA: Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* 2018;67:901–904.
38. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P: Spred3: Interactive visualization of spatio-temporal history and trait evolutionary processes. *Mol Biol Evol* 2016;33:2167–2169.
39. Hemelaar J: The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 2012;18:182–192.
40. Digban TO, Iweriebor BC, Obi LC, Nwodo U, Okoh AI: Analyses of HIV-1 Integrase gene Sequences among treatment naive patients in the Eastern Cape, South Africa. *J Med Virol* 2020;92:1165–1172.
41. Sivay MV, Hudelson SE, Wang J, *et al.*: HIV-1 diversity among young women in rural South Africa: HPTN 068. *PLoS One* 2018;13:e0198999.
42. Karade SK, Ghate MV, Chaturbhui DN, *et al.*: Cross-sectional study of virological failure and multinucleoside reverse transcriptase inhibitor resistance at 12 months of antiretroviral therapy in Western India. *Medicine (Baltimore)* 2016;95:e4886.
43. Sharma AL, Singh TR, Singh LS: Antiretroviral resistance, genotypic characterization, and origin of Human Immunodeficiency Virus among the infected wives of Intravenous drug users in Manipur. *Sci Rep* 2018;8:15183.
44. Corado ALG, Bello G, Leão RAC, Granja F, Naveca FG: HIV-1 genetic diversity and antiretroviral drug resistance among individuals from Roraima state, northern Brazil. *PLoS One* 2017;12:e0173894.
45. Idriss H, Kawa S, Damuni Z, Thompsona B, Wilson SH: HIV-1 reverse transcriptase is phosphorylated in vitro and in a cellular system. *Int J Biochem Cell Biol* 1999;31:1443–1452.
46. Ohtsukia K, Maekawaa T, Haradaa S, Karinoa A, Morikawab Y, Itoc M: Biochemical characterization of HIV-1 Rev as a potent activator of casein kinase II in vitro. *FEBS Lett* 1998;428:235–240.
47. Soares MA, Oliveira T, Brindeiro RM, *et al.*: A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. *AIDS* 2003;17:11–21.
48. Monteiro JP, Ferraro GA, Oliveira T, *et al.*: Genetic and biologic characterization of HIV type 1 subtype C isolates from South Brazil. *AIDS Res Hum Retroviruses* 2007;23:135–143.
49. Graf T, Vrancken B, Junqueira DM, *et al.*: Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *J Virol* 2015;89:12341–12348.
50. Silveira AA, Cardoso LPV, Francisco RBL, Stefani MMA: HIV type 1 molecular epidemiology in pol and GP genes among naive patients from Mato Grosso do Sul state, Central Western Brazil. *AIDS Res Hum Retroviruses* 2012;28:304–307.
51. Brígido LFM, Ferreira JLP, Almeida VC, *et al.*: Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. *AIDS Res Hum Retroviruses* 2011;27:339–344.
52. Brazil, Ministry do Tourism. Consumer survey travel intent. Getulio Vargas Foundation, São Paulo, 2016.
53. Brazil, Ministry do Tourism. Characterization and dimensioning of domestic tourism in Brazil. Economic Research Institute Foundation, São Paulo, 2012.

Address correspondence to:

Joana Paixão Monteiro Cunha
Núcleo de Bioinformática
Departamento de Bioquímica e Biofísica
Universidade Federal da Bahia
Av. Reitor Miguel Calmon, s/n
Salvador 40110902
Bahia
Brasil

E-mail: joana.monteiro@ufba.br