ESCOLA BAHIANA DE MEDICINA E SAÚDE PÚBLICA
CURSO BIOMEDICINA

GABRIEL ALVES DE CARVALHO

ORIGEM E DISPERSÃO DO SUBTIPO F1 DO HIV-1 NA BAHIA

SALVADOR
2024

GABRIEL ALVES DE CARVALHO

ORIGEM E DISPERSÃO DO SUBTIPO F1 DO HIV-1 NA BAHIA

Trabalho de Conclusão de Curso apresentado à Escola Bahiana de Medicina e Saúde Pública, como parte dos requisitos para obtenção do título de Bacharel em Biomedicina.

Orientador: Prof. Dra. Luciane Amorim Santos

SALVADOR

2024

GABRIEL ALVES DE CARVALHO

ORIGEM E DISPERSÃO DO SUBTIPO F1 DO HIV-1 NA BAHIA

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do grau de Bacharel em Biomedicina e aprovada em sua forma final pelo Curso de Biomedicina da Escola Bahiana de Medicina e Saúde Pública.

Salvador, 6 de novembro de 2024.

Prof. Título. Nome
LOCAL DE ATUAÇÃO

Prof. Título. Nome
LOCAL DE ATUAÇÃO

Prof. Título. Nome
LOCAL DE ATUAÇÃO

AGRADECIMENTOS

A realização deste trabalho não seria possível sem o apoio e a contribuição de várias pessoas, às quais sou imensamente grato.

Agradeço primeiramente aos meus pais, pelo apoio incondicional em todos os momentos da minha vida acadêmica.

Aos meus professores e orientadores, que com paciência e sabedoria me guiaram ao longo dos anos de inciação científica. Especialmente ao professor Ricardo Khouri e orientadora Luciane Amorim.

Agradeço também aos Profs. Marta Giovanetti e Dennis pelo tempo dedicado e contribuições valiosas.

Aos meus colegas da Fiocruz Bahia e da Escola Bahiana de Medicina, pela troca de conhecimentos, pelas conversas e apoio mútuo.

Por fim, agradeço a todos aqueles que, direta ou indiretamente, contribuíram para a conclusão deste TCC.

“Uma coisa eu aprendi: que toda a nossa ciência, comparada com a realidade, é primitiva e infantil - e ainda assim é a coisa mais preciosa que temos.”

Albert Einstein

RESUMO

De 2007 a 2023, o Brasil relatou 489,594 casos de infecção pelo HIV-1, com a região Nordeste sendo a segunda região mais afetada. Dentro do Nordeste, a Bahia, como o estado mais populoso, detém o maior número absoluto de casos. Embora o subtipo B prevaleça no Nordeste, variantes não B, como F1, representam uma proporção significativa, necessitando de análises filogenéticas mais aprofundadas. Este estudo utilizou dados genéticos e epidemiológicos para explorar a origem e a disseminação do subtipo F1 do HIV-1 na Bahia. Analisamos 24 novas sequências do gene *pol* do HIV-1 de pessoas vivendo com vírus no estado da Bahia, Brasil. Para entender a filodinâmica viral, integramos sequências do gene *pol* do banco de dados de Los Alamos, resultando em um conjunto de dados de 5195 sequências. O achado principal das análises filodinâmicas e filogeográficas revela múltiplas introduções na Bahia entre a década de 1990 e o início dos anos 2000, com a região Sudeste-BR (estados de São Paulo, Rio de Janeiro e Minas Gerais) sendo a principal fonte de introduções na Bahia. Além disso, notamos que clados mais recentes (2010–2017) de Pernambuco e do estado do Pará têm alta similaridade com o principal clado da Bahia identificado, o que indica uma possível origem ancestral comum compartilhada dessas epidemias localizadas. Em conclusão, é muito importante continuar a aquisição de novas sequências do HIV-1 para monitorar em tempo real a dinâmica espaço-temporal dos subtipos circulantes, para que em um futuro próximo as estratégias de vacinação e novos tratamentos possam ser aplicados de forma efetiva.

Palavras-chave: HIV-1; Subtipo F1; Filogenética; Filodinâmica;

ABSTRACT

From 2007 to 2023, Brazil reported 489,594 cases of HIV-1 infection, with the Northeast region ranking as the second most affected area. Within the Northeast, Bahia, as the most populous state, holds the highest absolute number of cases. Although subtype B prevails in the Northeast, non-B variants, like F1, represent a significant proportion, needing comprehensive phylogenetic analyses. This study employed genetic and epidemiological data to explore the origin and spread of HIV-1 subtype F1 in Bahia. We analyzed 24 new gene *pol* sequences from people living with HIV from the state of Bahia, Brazil. To understand viral phylodynamics, we integrated sequences from the *pol* region from the Los Alamos database, resulting in a dataset of 5195 sequences. The primary finding of the phylodynamic and phylogeography analyses revealed multiple introductions in Bahia between the early 1990s to the 2000s, with the Southeast-BR region (states of Sao Paulo, Rio de Janeiro and Minas Gerais) being the main source of introductions to Bahia. Furthermore, we noted that more recent clades (2010– 2017) from Pernambuco and the Pará state have very high similarity with the main Bahia clade identified, which indicates for a possible shared common ancestor origin of those localized epidemics. In conclusion, it is very important to continue the obtention of novel HIV-1 sequences for monitoring in real-time the spatiotemporal dynamics of the circulating subtypes, so in a short future vaccination strategies and novel treatment could be more precise.

Keywords: HIV-1; F1 subtype; Phylogenetic; Phylodynamics.

SUMÁRIO

viruses

MDPI

Origin and dispersal of HIV-1 subtype F1 in Bahia, Northeast Brazil

Gabriel Carvalho[1,2], Laise de Moraes[2,3], Daniel Polita[6], Thayná Gonzalez[2], Filipe Ferreira de Almeida Rego[1], Dennis Maletich Junqueira[6], Marta Giovanetti[5], Ricardo Khouri[2,3,4] and Luciane Amorim Santos[1,2,3]*

[1] Escola Bahiana de Medicina e Saúde Pública, Avenida Dom João VI, 275, Brotas, Salvador 40290-000, Bahia, Brazil.
[2] Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Candeal, Salvador 40296-710, Bahia, Brazil.
[3] Programa de Pós-Graduação em Ciências da Saúde, Faculdade de Medicina da Bahia, Universidade Federalda Bahia, Praça Ramos de Queirós, s/n, Largo do Terreiro de Jesus, Salvador 40026-010, Bahia, Brazil.
[4] Rega Institute for Medical Research, Department of Immunology, Microbiology and Transplantation, KU Leuven, Herestraat 49-Box 1030, 3000 Leuven, Belgium.
[5] Sciences and Technologies for Sustainable Development and One Health, University of Campus Bio-Medico, 21-00128 Rome, Italy.
[6] Laboratório de Bioinformática e Evolução de Vírus, Departamento de Bioquímica e Biologia Molecular, Centro de Ciências Naturais e Exatas (CCNE), Universidade Federal de Santa Maria (UFSM), Santa Maria, Brazil.

* Correspondence: luciane.tika@gmail.com; Tel.: +55-71-991690271

Abstract: From 2007 to 2021, Brazil reported 381,793 cases of HIV-1 infection, with the Northeast region ranking as the second most affected area. Within the Northeast, Bahia, as the most populous state, holds the highest absolute number of cases. Although subtype B prevails in the Northeast, non-B variants, like F1, represent a significant proportion, needing comprehensive phylogenetic analyses. This study employed genetic and epidemiological data to explore the origin and spread of HIV-1 subtype F1 in Bahia. We analyzed 24 new gene *pol* sequences from people living with HIV from the state of Bahia, Brazil. To understand viral phylodynamics, we integrated sequences from the pol region from the Los Alamos database, resulting in a dataset of 5195 sequences. The primary finding of the Maximum Clade Credibility (MCC) trees revealed multiple introductions in Bahia between the early 1990s to the 2000s, with the Southeast-BR region (states of Sao Paulo, Rio de Janeiro and Minas Gerais) being the main source of introductions to Bahia. Furthermore, we noted that more recent clades (2010– 2017) from Pernambuco and the Pará state have very high similarity with the main Bahia clade identified, which indicates for a possible shared common ancestor origin of those localized epidemics. In conclusion, it is very important to continue the obtention of novel HIV-1 sequences for monitoring in real-time the spatiotemporal dynamics of the circulating subtypes, so in a short future vaccination strategies and novel treatment could be more precise.

Keywords: HIV-1; F1 subtype; Phylogenetic; Phylodynamics.

1. Introduction

The human immunodeficiency virus (HIV) can lead to acquired immunodeficiency syndrome (AIDS), affecting CD4+ T lymphocytes and causing severe immunological deficiency. Despite the advances in antiretroviral therapy (ART) leading to viral load suppression and control of disease progression, some cases of HIV can be resistant to treatment [1]. In 2023, UNAIDS estimated that around 39.9 million people were living with HIV worldwide, with approximately 60% of these infections in sub-Saharan Africa [2].

AIDS is a global pandemic caused by two genetically distinct lentiviruses, HIV-1 and HIV-2, which originated from multiple cross-species transmissions of simian immunodeficient viruses (SIV) from non-human primates to humans. HIV-1 has multiple groups M (Major), O (Outlier), N and the more recent group P. Group M is predominant worldwide and has been the major driver of the global HIV pandemic. In contrast, groups N, O, and P have remained less widespread, with cases identified mailing in Africa. Group M viruses are further divided into subtypes (A-D, F-H, J, K) and recombinant forms [3].

From 2007 to 2023, 489,594 new cases of HIV-1 infection were registered in Brazil, with the Northeast region being the second most affected, and Bahia, as the most populous state in the region, occupying the second position in absolute number of regional cases [4]. It is known that the distribution of viral subtypes around the world, the severity of the disease, and the risk groups associated with the subtype are different. For these reasons, real-time monitoring of the dynamics of each viral subtype is important to help in public health policy and clinical conduct.

In Bahia, the most prevalent subtype is B (68.1%), followed by the recombinant BF (18.6%) and subtype F1 (8.7%) [5]. According to previews studies in molecular epidemiology, the F1 subtype of the HIV-1 virus was introduced in Brazil around 1977 and gradually spread throughout the population, reaching more stable levels in the 1990s [6]. Despite this, no study has yet described the molecular epidemiology and dissemination patterns of subtype F1 in Bahia. Therefore, it is crucial to continue the genomic surveillance and the obtention of new sequences to track the evolutionary dynamics and spread of each subtype within the study region.

As new sequences are generated and made available in public databases, it becomes possible to construct a highly diverse dataset that can address questions related to molecular epidemiology in Bahia and other Northeast states, where there has been a limited number of studies and available data for analysis. This can be attributed to a lack of sequencing and genomic surveillance in past decades. In this context, the present study aimed to delve deeper into the dispersion and origin of the F1 subtype in the state of Bahia. Here we evaluated 24 new sequences of the HIV-1 *pol* gene from Bahia, integrating with a large dataset of global sequences to perform phylogenetic, phylodynamic, and phylogeographic analyses.

2. Materials and Methods

2.1 HIV-1 Samples from Bahia

In this study, the unknown origin and dispersion of HIV-1 F1 subtype in the state of Bahia, Northeast Brazil were explored. To achieve this, 24 new *pol* gene sequences classified as belonging to the F1 subtype which were obtained in a previous study at the Precision Medicine and Public Health Laboratory (MESP²), Instituto Gonçalo Moniz (IGM/FIOCRUZ) were analyzed. The sequences were obtained using the Sanger methodology, from samples of people living with HIV (PLWH) monitored by the State Center Specialized in Diagnosis, Assistance and Research (CEDAP), Bahia, Brazil, as previously described [6]. These are residents of the state of Bahia and were diagnosed with HIV between 2014 and 2016.

2.2 Dataset construction

In addition to the 24 HIV-1F1 Bahia sequences, a large and genetically diverse dataset for phylogenetic analysis were assembled by searching for worldwide HIV-1F1 sequences available in the public database. These sequences were obtained in the Los Alamos database (Available at: www.hiv.lanl.gov/components/sequence/HIV/search). The search was filtered following the selection criteria: belong to the F1 subtype; *pol* gene from position 2085 to 5086 according to HXB2 reference genome (accession number K03455.1); fragments with a minimum length of 700 base pairs (bp); with collection date; and location information. The initial dataset comprised 5195 sequences. Thereafter, duplicate sequences and/or those without metadata were removed. Finally, all information

including accession number, sample location, date of collection and nucleotide sequence were organized in a in Microsoft Excel 2021 spreadsheet. To ensure that all sequences were HIV-1 subtype F1 and no recombinants were included in the analyses, all sequences were submitted to the Genome Detective platform, accessing the REGA HIV-1 subtyping tool version 3.46 [7]. All recombinant F1 sequences were removed.

2.3 Phylogenetic and transmission clusters identification

Next, the sequence alignment was performed using the HXB2 genomic reference as a guide, through the MAFFT (Multiple Alignment using Fast Fourier Transform) [8] and edited using AliView [9]. The final edited alignment comprised 4493 *pol* sequences. From the final edited alignment, the initial phylogeny was constructed using the maximum likelihood (ML) method in IQ-TREE [10,11] under the GTRT+F+I+G4 evolutionary model selected by ModelFinder, with 1000 replicates of ultrafast bootstrap and approximate likelihood-ratio test (aLRT – approximate maximum likelihood ratio test) to evaluate the support of the branches. The Microreact platform and the FigTreeV1.4.4 program was used to visualize and plot the tree.

To extract more precise information from the maximum likelihood, we employed the Cluster Picker tool [12] an automated tool designed to detect monophyletic clades meeting the user-input criteria, including branch support values and genetic distance between taxa. We tested aLRT support values ranging from 80 to 95 values and genetic distances threshold from 0.5% to 6% to compare and identify the optimal cut-off for our analysis. Finally, we plotted the Cluster Picker output data on maps using 'ggplot2' package in RStudio.

2.4 Migration patterns analysis

To better understand the migratory flow of the HIV-1 subtype F1 across Brazilian regions and potential phylogenetic interactions with Bahia, first the molecular clock signal of the maximum likelihood tree was assessed using the root-to-tip regression method available at TempEst v1.5.2 [13] following the removal of potential outliers that may violate the molecular clock assumption. Subsequently, the time-scaled tree topology using TreeTime [14] was constructed and then followed up with discrete ancestral states reconstruction using the 'mugration' package which is an extension of TreeTime. Each location state change from one node to another was counted with a Python script called 'AncestralChanges' by iterating over the time-scaled phylogeny from the root to tip. Finally, the generated data about state changes across locations were plotted in RStudio, using the 'ggalluvial' R package.

2.5 Bayesian reconstruction and molecular dating of introductions to Bahia

To infer spatiotemporal spread of HIV-1F1 to Bahia state, two different clades containing the highest number of Bahia sequences were extracted directly from the maximum likelihood tree and were used in the analysis. Initially, temporal signal of each clade was assessed using TempEst v1.5.3 [13] and temporal structure was accepted if the correlation coefficient was ≥ 0.7. In BEAUti v1.10.4, the uncorrelated relaxed molecular clock and the Bayesian Skygrid model were selected. Moreover, discrete location traits were informed for ancestral states reconstruction. Maximum Clade Credibility trees were obtained after three independent runs of 100 million iterations with sampling every 10,000 steps in BEAST v.1.10.4. [15] The convergence of each run was assessed in Tracer v1.7.2 [16] Effective Sample Size (ESS) of all parameters must be ≥ 200. MCC trees from each run were combined in LogCombiner v1.10.4 and summarized in TreeAnnotator v1.10.4 after discarding the initial 10% burn-in [15]. Posterior Probability (PP) of the branches was considered significant when ≥ 0.9. Finally, the spatiotemporal history was inferred from the posterior trees and data was plotted on Brazil and South America maps using RStudio and a custom R script adapted from R "seraphim" package.

3. Results

The initial tree was obtained using the Maximum Likelihood (ML) method and has 4,493 sequences (Figure 1). In this tree the sequences were divided into two large clades, one composed of sequences mostly from Europe and Africa (F1.2), and the other with mostly Brazilian sequences and some isolated from Europe (F1.1). In the Brazilian clade, it was possible to observe multiple events of viral introduction to the state of Bahia that led to the establishment of different subtype F1 viruses from different transmission sources within the state.

To identify the migration patterns, the ML tree was transformed into a time-scaled tree using the TreeTime tool. The location transitions between the ancestral nodes of the tree were counted using the logic of "Origin" and "Destination" (Figure 1B). The largest HIV-1F1 flow observed is from the Southeast of Brazil (SoutheastBR) as "Origin" spreading to all other Brazilian regions, including Bahia, and other parts of the world, indicating the important role of this region in the HIV-1F1 viral dissemination.

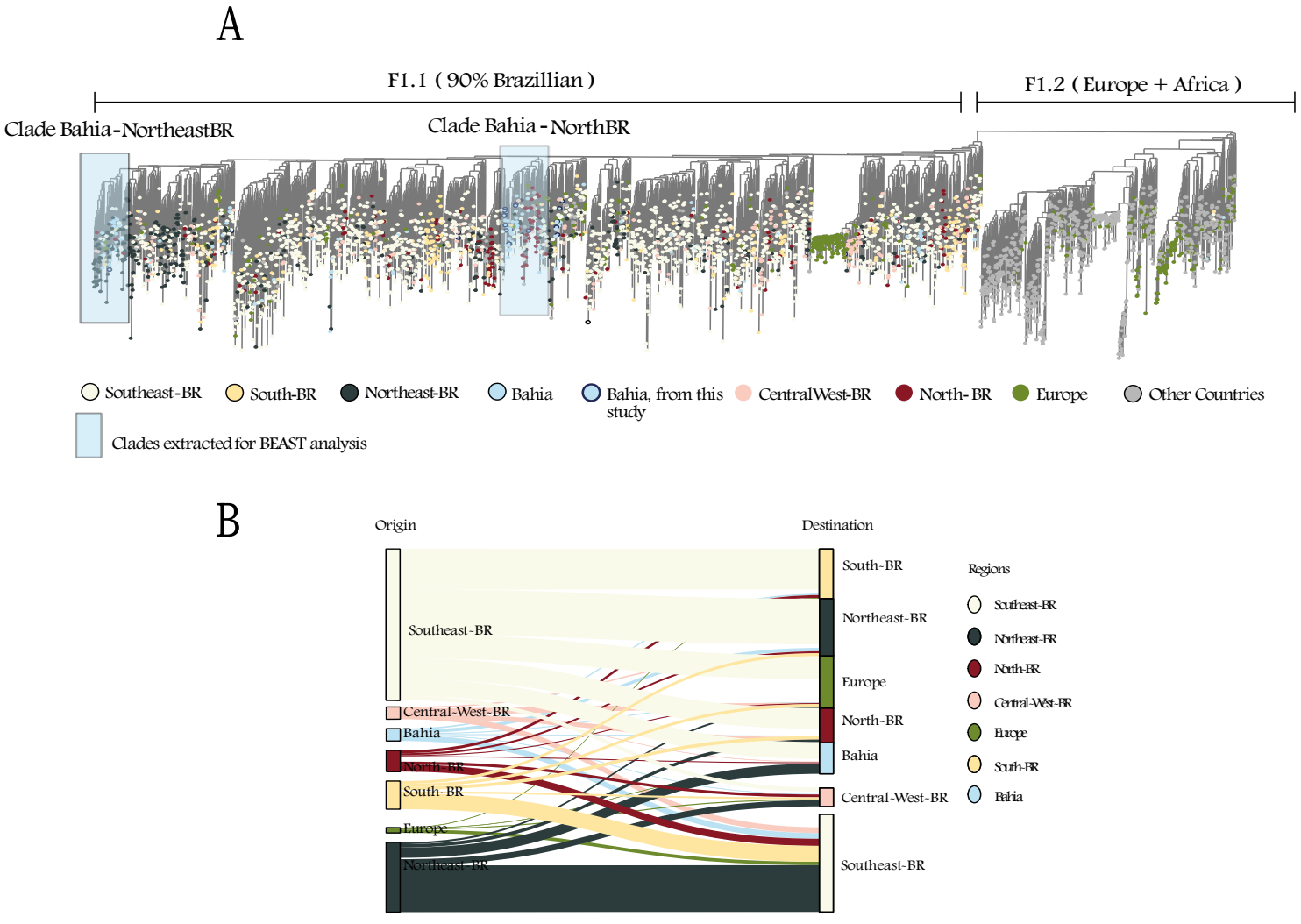

Figure 1. (a) Maximum Likelihood tree containing 4493 gene *pol* sequences of HIV-1 subtype F1; **(b)** HIV-1 F1 migration flow between Brazilian regions, generated from the initial maximum likelihood tree (4993 seqs) by time-scaling in TreeTime tool and 'mugration' package. Each color represents one specific region from BR (Brazil).

To investigate the evolutionary dynamics of HIV-1F1 in the state of Bahia using a Bayesian reconstruction, the two clades with the greatest number of sequences from the state were extracted from the ML tree: Bahia-NorthBR and Bahia-NortheastBR. The Bahia-NorthBR clade (Figure 2A), with 147 sequences from 1996 to 2017, was mainly represented by Bahia (45) and the NorthBR region (35); and the Bahia-NortheastBR clade (Figure 3A),

consisting of 105 sequences from 1998 to 2017, primarily included sequences from Bahia (28) and Pernambuco (45).

The tMRCA of the Bahia-NorthBR clade was from 1985 with a 95 % HPD interval of [1976;1991]. The tree root geographic origin was inferred as the Southeast region of Brazil (Figure 2A). When observing the root of this clade, it is possible to observe that it branches into two large groups of sequences, one composed mainly of the NorthBR region and the other with sequences from Bahia. This demonstrates the common ancestry of both introductions that may have been originating from the same virus that was circulating in Southeast of Brazil. However, the two transmission chains developed and evolved independently despite the genetic similarity of the clades. The Southeast-BR route to Bahia is very evident and several ancestral states transitions were detected in the tree. Transitions can be considered transmission events that go from one region to another and the ones to Bahia are shown in (Table 1). Six occurrences with "Origin" in Southeast-BR and "Destination" in Bahia were identified in the Bahia-NorthBR clade. The first events occurred in 1998, HPD 95% [1995;2001] and 1999 with HPD 95% [1996;2001]. Additionally, the transitions involving the other regions in the tree are detailed in **(**Figures 2B-D**)**.

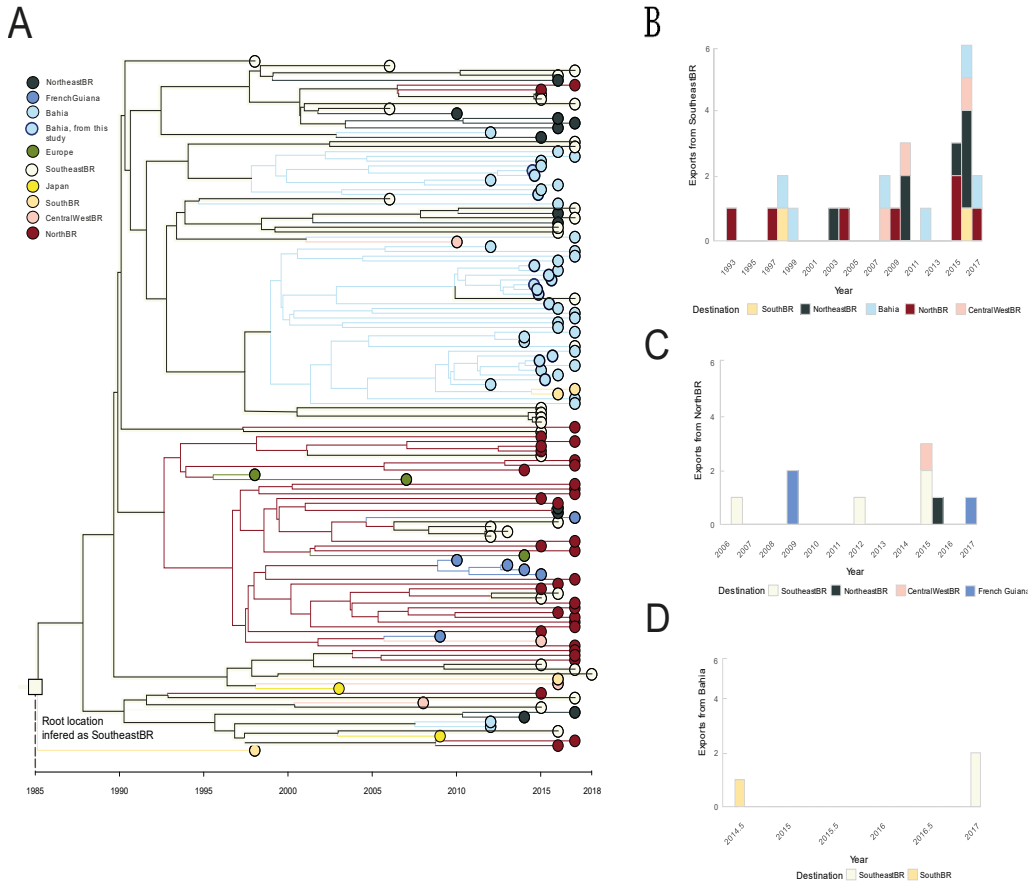

Figure 2. (a) Maximum Clade Credibility Tree for the Clade "Bahia-NorthBR". Ancestral states are represented by colored branches corresponding to their geographical locations; **(b,c,d**) Transitions of the subtype F1 virus from one location to another, with each location colored as it appears in the MCC tree.

Regarding the second clade extracted from ML for Bayesian reconstruction, Bahia-NortheastBR (Figure 3A), tMRCA of tree root was 1987 with HPD 95% [1982;1992] Although we have a sequence from Southeast-BR in more basal positions of the clade, it was not possible to obtain statistical support that this clade would also originate from the Southeast region of Brazil. However, it was possible to identify that the sequences from Bahia have high similarity with other sequences identified in the tree as Northeast-BR. These sequences mostly belong to the state of Pernambuco. Some transition events with

"Origin" in Bahia and "Destination" in Northeast-BR were identified and are shown in (Figure 3B-C). Furthermore, both introductions to Bahia detected in the Bahia-NortheastBR clade have a tMRCA of 1992 with a 95% HPD [1989; 1996], and a tMRCA of 1995 with a 95% HPD [1989; 1999]. Despite this, neither could be traced back to their origins due to a lack of older sequences in the clade that would resolve the long branches and clarify the origin of these specific introductions.

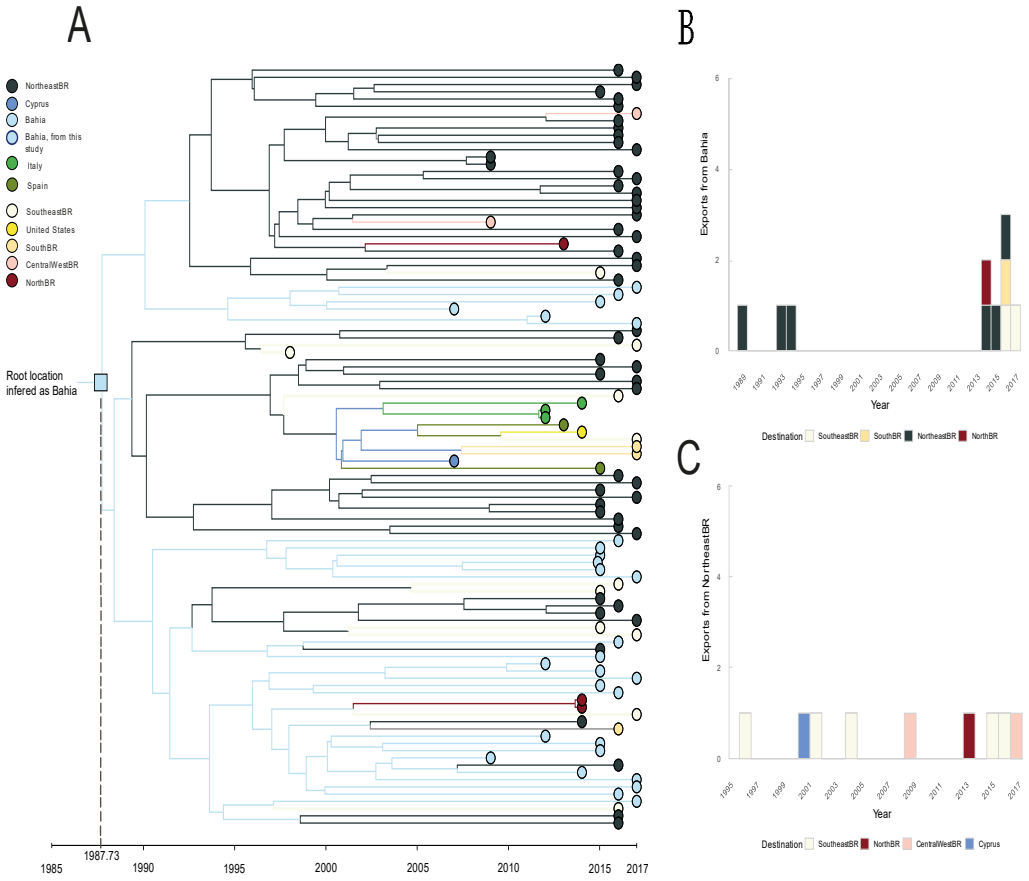

Figure 3 (a) Maximum Clade Credibility Tree for the Clade "Bahia-NortheastBR." Ancestral states are represented by colored branches corresponding to their geographical location; **(b,c)** Transitions of the subtype F1 virus from one location to another, with each location colored as it appears in the MCC tree "BahiaNortheast-BR"

Table 1. Characterization of the HIV-1 F1 discrete ancestral state transitions to Bahia state obtained from MCC trees. Date of occurrence represents the mean date estimated for the transitions among locations. The origin is the most possible source of the event.

Clade	Date of transition occurrence	HPD 95 %	Origin/Destination
Clade BahiaNortheastBR	1992	[1989;1996]	Unknown to Bahia
	1995	[1989;1999]	Unknown to Bahia
Clade BahiaNorthBR	1998	[1995; 2001]	SoutheastBR to Bahia
	1999	[1996; 2001]	SoutheastBR to Bahia
	2008	[2004; 2010]	SoutheastBR to Bahia
	2012	[2007; 2012]	SoutheastBR to Bahia
	2016	[2010; 2016]	SoutheastBR to Bahia
	2017	[2012; 2017]	SoutheastBR to Bahia

Additionally, cluster analysis (Figure 4) showed a significant intrastate concentration of transmission chains, with a few cases extending beyond state boundaries. Although Bahia's main source of transmission is the Southeast, with introductions dating back to the 1990s, cluster analysis shows that the transmission of the HIV-1 subtype F1 is generally occurring more locally within respective regions. The map in (Figure 4A) illustrates that out of 240 clusters identified in Brazil, 87.5% are intrastate, while only 12.5% are interstate (involving two or more states).

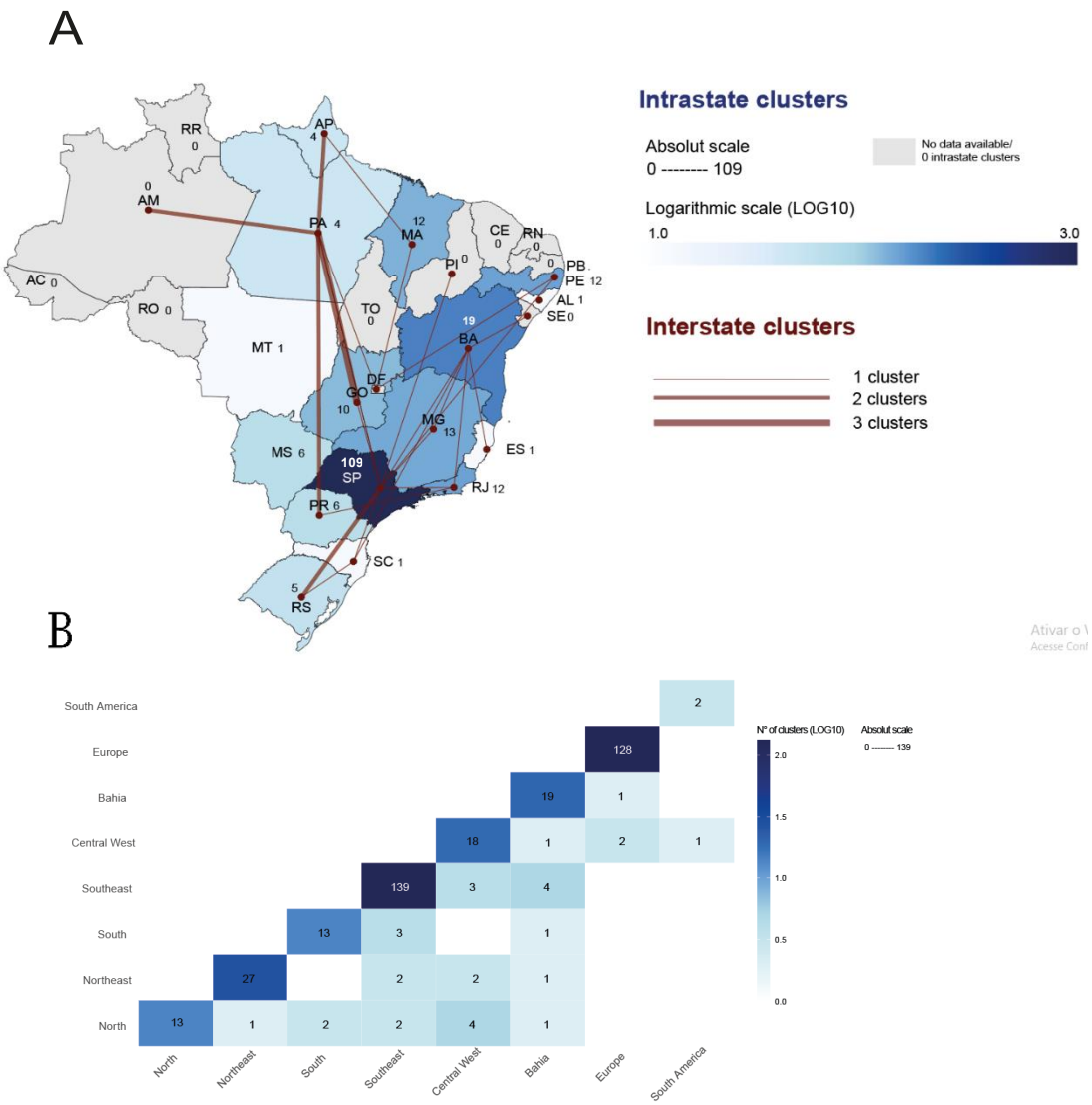

Figure 4 (a) Number of intrastate and interstate clusters distributed by Brazilian states (UF); **(b)** Heatmap with the number of pairwise clusters by local.

The heatmap in (Figure 4B) presents the same data in a different format, showing the number of clusters formed by pairs of location and emphasizing the strong concentration of intrastate clusters. Bahia is involved in 27 clusters, 19 of which are intrastate and 9 interstates. Of the 9 interstate clusters, 4 involve Bahia and Southeast-BR, while the remaining 5 involve Bahia and other regions of Brazil.

4. Discussion

This study employed genetic and epidemiological data to infer the origin and dispersion of the HIV-1 F1 subtype in Bahia and the possible correlations between

Brazilian regions. Here, 24 new HIV-1 *pol* gene sequences were analyzed and contributed to the ongoing efforts in genomic surveillance and molecular epidemiology. It provides valuable insights that can enhance decision-making in public health by deepening our understanding of the virus's dispersion patterns in Bahia and Brazil. The main findings reveal that the Southeast region of Brazil was the primary source of HIV-1 introductions to Bahia and other Brazilian states.

From the Maximum Likelihood tree, it was possible to explicitly identify the segregation of the F1 subtype into two large clades. In a previous study, the same clustering pattern was found, in which sequences from Brazil and Europe, more specifically Spain and Italy, could be classified as a sub-subtype of F1 (F1.1) clustered in monophyletic clade. In contrast, the other large monophyletic clade composed of sequences from Africa and Romania could be classified in F1.2 [17]. Initially, it is noted that within this large clade composed mainly of Brazil and Europe (F1.1), there are still some genetically more distant groups from each other, which leads to the hypothesis that although the subtype in Brazil has a single-entry origin, several viruses within the subtype established themselves independently in various regions. In the most basal positions in the ML tree, many sequences from Southeast Brazil and Europe are noted, bringing to light the previously mentioned hypothesis that the F1 subtype arrived in Brazil through European importation. These findings are similar to those found for other subtypes, such as D and C, which also suggest Europe as the source of HIV-1 introductions to Brazil [18,6].

After arriving in Brazil, the F1 subtype spread to São Paulo, Rio de Janeiro, and Minas Gerais with greater intensity, and later established itself in states in the North (Pará and Amapá), Central West (Goias), and states in the Northeast, mainly Bahia and Pernambuco. Specifically in the Northeast region, multiple independent introduction events of F1 occurred; however, most circulating viruses of this subtype in Brazil have high genetic similarity and might originated from a common ancestor and introduction. A similar dissemination pattern was found in a previous study [19], which elucidated that introductions of subtype C into the Brazilian Northeast likely occurred between 1980 and 1990 and originated from the South region of the country, where the virus circulation began in the 1970s, making it the region with the highest prevalence of subtype C. Currently, although this subtype has been detected across most of Brazil, it is still more concentrated in the South. The difference between the spreading of subtype C and F1 is that F1 managed to establish itself successfully in several regions of Brazil despite being initially concentrated in the Southeast. Today, it has a high prevalence rate in the Northeast, together with subtype B and the BF recombinant [20].

Unfortunately, it is not possible to identify all movements of the subtype F1 considering all Brazilian states, since some of them do not yet have genomes available in sufficient quantity. Bahia has more than 100 genomes available in addition to the 24 new ones generated in this study, which made it possible to identify several Bahia clades in the ML. It is possible to consider each clade as a different introduction that occurred at a given time with its respective confidence interval, as well as its most likely geographic origin. The two main clades from Bahia that underwent Bayesian reconstruction and phylogeography provide important insights and point to the Southeast region of Brazil as the primary source of introductions. However, it is very likely that, due to the lack of significant HIV-1 virus sequencing in the North and Northeast regions, the connections found between these regions and Bahia may be underestimated by phylogenetic analyses in general.

Finally, the analysis of HIV-1F1 transmission “clusters” or “chains” shows that the identified clusters are mainly formed by sequences from the same Brazilian region or state. This demonstrates that the occurrence of new introductions between different regions is becoming rarer, a fact that can be related to the success of Antiretroviral Therapies (ART) implemented in Brazil. The transmissions that still occur are increasingly restricted to small groups of people, and some isolated cases of interstate clusters have been identified. As previously mentioned, Cluster Picker is a tool that selects

monophyletic clades in the ML tree and allows the user to adjust the genetic distance parameters between sequences and the branch support, such as aLRT. This approach has already been used in previous studies [11,21] and it was successful in detecting samples linked to HIV epidemics in Europe. For the Brazilian scenario, it was also possible to extract, from the selected clusters identified by the tool, which sequences are more likely to be part of the same transmission chain and epidemiologically linked

This study presents the first molecular epidemiology analysis of subtype F1 sequences in the state of Bahia and describes transmission patterns between Brazilian states over time using approaches such as phylodynamic and phylogeography. It is important to emphasize that the reliability of the results from this type of analysis is highly dependent on effective genomic surveillance in the country, which, although it has improved over the last decade with the development of the National HIV-1 Genotyping Network (RENAGENO), remains imbalanced in terms of the number of sequences available in databases for each Brazilian region. This can be explained by the lack of resources and research investment in the regions with lower GDP in the country (North and Northeast). Therefore, continued genomic surveillance with greater intensity in less advantaged regions is necessary to reduce phylogenetic bias in analyses. Additionally, the genotyping of samples from individuals infected with HIV-1 is important not only for molecular epidemiology analyses but also for clinical management and better selection of ART (Antiretroviral Therapies).

5. Conclusions

In conclusion, this study brings that most subtype F1 clusters are composed of sequences from the same regions. This suggests that new introductions across different regions have become less common, possibly due to the effectiveness of Antiretroviral Therapies (ART). Notably, the study indicates that introductions to Bahia from the Southeast of Brazil occurred between the early 1990s to 2000s. However, the lack of genetic data from the North and Northeast regions highlights the need for better genomic surveillance in these areas. Addressing this gap is essential to ensure accurate analysis of HIV-1 epidemics. Additionally, testing samples from infected individuals is important not just for understanding the epidemic but also for improving treatment choices. Overall, this study highlights the importance of ongoing genomic research and monitoring to effectively address HIV-1 in Brazil.

Author Contributions: Conceptualization, R.K.; L.A.S. and M.G..; data curation, F.F.d.A.R.; L.d.M., G.A.d.C. and T.T.G.; formal analysis, G.A.d.C. and L.A.S., methodology, G.A.d.C ; R.K.; L.A.S.; M.G.; D.M.J. and D.P., resources, R.K and M.G..; writing—original draft, G.A.d.C and L.A.S. writing—review and editing, L.A.S.; R.K; F.F.d.A.R.; L.d.M and M.G.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) [Finance Code 001, R.K.], Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, grant APP0032/2016, R.K.), Brazilian National Council for Scientific and Technological Development (CNPq, grant 65083/2015-8, L.A.S.). MG is funded by PON "Ricerca e Innovazione" 2014–2020. None of the funding organizations had any role in the study design, data collection, data interpretation, or writing of this report.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the Gonçalo Moniz Institute (IGMFIOCRUZ) (CEP/CAAE: 51733115.1.0000.0040; approval number: 1.764.505).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study and the use of serum samples used in this research were collected for diagnostic purposes. These samples were accompanied by their respective epidemiological sheets, with patient identification already encoded.

Data Availability Statement: All data, scripts, and new sequences generated will be made available at https://github.com/gabalves1. New sequences IDs: (HV0001, HV0003, HV0014, HV0019, HV0020,

HV0022, HV0024, HV0064, HV0032, HV0079, HV0084, HV0092, HV0105, HV0116, HV0119, HV0151, HV0199, HV0248, HV0348, HV0354, HV0357, HV0406, HV0423, HV0444).

Acknowledgments: We acknowledge the people living with HIV-1 from the CEDAP for their consent and enrolment in this study.

Conflicts of Interest: The authors declare that there are no conflicts of interest related to this study.

References

1. Fanales-Belasio, E.; Raimondo, M.; Suligoi, B.; Buttò, S. HIV virology and pathogenetic mechanisms of infection: a brief overview. Ann. Ist. Super. Sanità. **2010**, 46, 1.
2. UNAIDS. Global AIDS update 2023: data book. Available online: https://www.unaids.org/sites/default/files/media_asset/data-book-2023_en.pdf (accessed on 26 October **2024**).
3. Sharp, P.M.; Hahn, B.H. Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. **2011**, 1, a006841.
4. Ministério da Saúde. Boletim Epidemiológico - HIV e Aids 2023. Available online: https://www.gov.br/aids/pt-br/central-de-conteudo/boletins-epidemiologicos/2023/hiv-aids/boletim-epidemiologico-hiv-e-aids-2023.pdf/view (accessed on 26 October **2024**).
5. Amaral, A.G.M.; Oliveira, I.B.; Carneiro, D.C.; Alcantara, L.C.J.; Monteiro-Cunha, J.P. An overview of the molecular and epidemiological features of HIV-1 infection in two major cities of Bahia state, Brazil. Mem. Inst. Oswaldo Cruz. **2017**, 112, 411–418.
6. de Almeida Rego, F.F.; de Moraes, L.; Giovanetti, M.; Silva, J.A.G.; Torres, F.G.; de Oliveira Silva, M.; Khouri, R. Genomic detection of the emerging, highly pathogenic HIV-1 subtype D in Bahia, Northeast Brazil. Viruses. **2023**, 15, 8.
7. Pineda-Peña, A.-C.; Faria, N.R.; Imbrechts, S.; Libin, P.; Abecasis, A.B.; Deforche, K.; Vandamme, A.-M. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Infect. Genet. Evol. **2013**, 19, 337–348.
8. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. **2019**, 20, 1160–1166.
9. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. **2014**, 30, 3276–3278.
10. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **2015**, 32, 268–274.
11. Hoang, D. T.; Chernomor, O.; von Haeseler, A.; Minh, B. Q.; Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. **2018**, 35(2), 518–522.
12. Ragonnet-Cronin, M.; Hodcroft, E.; Hué, S.; Fearnhill, E.; Delpech, V.; Brown, A.J.L.; Lycett, S. Automated analysis of phylogenetic clusters. BMC Bioinformatics. **2013**, 14, 317.
13. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. **2016**, 2, vew007.
14. Sagulenko, P.; Puller, V.; Neher, R.A. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. **2018**, 4, vex042.
15. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **2007**, 7, 214.
16. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst. Biol. **2018**, 67, 901–904.
17. Silva, G. P. S.; Oliveira, R. C.; de Souza, J. S. M.; Giovanetti, M.; Guimarães, M. L.; Brites, C.; Monteiro-Cunha, J. Tracing the relationship among HIV-1 sub-subtype F1 strains: a phylodynamic perspective. Mem. Inst. Oswaldo Cruz **2023**, 117.
18. Gräf, T.; Pinto, A.R. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. Virology. **2012**, 435, 170–178.
19. Oliveira, R. C.; Gräf, T.; Rego, F. F. de A.; Silva, G. P. S. A.; Giovanetti, M.; Monteiro Cunha, J. P. Dynamic dispersion of HIV-1 subtype C toward Brazilian Northeastern region. AIDS Res. Hum. Retroviruses. **2021**, 37(12), 913–921
20. Delatorre, E.; Couto-Fernandez, J.C.; Bello, G. HIV-1 genetic diversity in northeastern Brazil: High prevalence of non-B subtypes. AIDS Res. Hum. Retroviruses. **2017**, 33, 639–647.
21. Rose, R.; Lamers, S.L.; Dollar, J.J.; Grabowski, M.K.; Hodcroft, E.B.; Ragonnet-Cronin, M.; Laeyendecker, O. Identifying transmission clusters with cluster picker and HIV-TRACE. AIDS Res. Hum. Retroviruses. **2017**, 33, 211–218.

REVISTA CIENTÍFICA:
Link das regras para autores: https://www.mdpi.com/journal/viruses/instructions

REGRAS PARA OS AUTORES

1) Submission Process

Manuscripts for Viruses should be submitted online at susy.mdpi.com. The submitting author, who is generally the corresponding author, is responsible for the manuscript during the submission and peer review process. The submitting author must ensure that all eligible co-authors have been included in the author list (read the criteria to qualify for authorship) and that they have all read and approved the submitted version of the manuscript. To submit your manuscript, register and log in to the submission website. Once you have registered, click here to go to the submission form for Viruses. All co-authors can see the manuscript details in the submission system, if they register and log in using the e-mail address provided during manuscript submission.

2) Accepted File Formats

Authors are encouraged to use the Microsoft Word template or LaTeX template to prepare their manuscript. Using the template file will substantially shorten the time to complete copy-editing and publication of accepted manuscripts. The total amount of data for all files must not exceed 120 MB. If this is a problem, please contact the Editorial Office viruses@mdpi.com. Accepted file formats are:

- Microsoft Word: Manuscripts prepared in Microsoft Word must be converted into a single file before submission. When preparing manuscripts in Microsoft Word, we encourage you to use the Viruses Microsoft Word template file. Please insert your graphics (schemes, figures, etc.) in the main text after the paragraph of its first citation.
- LaTeX: Manuscripts prepared in LaTeX must be collated into one ZIP folder (including all source files and images, so that the Editorial Office can recompile the submitted PDF). When preparing manuscripts in LaTeX, we encourage you to use the Viruses LaTeX template files. You can now also use the online application writeLaTeX to submit articles directly to Viruses. The MDPI LaTeX template file should be selected from the writeLaTeX template gallery.

3) Supplementary files: May be any format, but it is recommended that you use common, non-proprietary formats where possible (see below for further details).
Disclaimer: Usage of these templates is exclusively intended for submission to the journal for peer review, and strictly limited to this purpose and it cannot be used for posting online on preprint servers or other websites.

4) Free Format Submission

Viruses now accepts free format submission:

- We do not have strict formatting requirements, but all manuscripts must contain the required sections: Author Information, Abstract, Keywords, Introduction, Materials & Methods, Results, Conclusions, Figures and Tables with Captions, Funding Information, Author Contributions, Conflict of Interest and other Ethics Statements. Check the Journal Instructions for Authors for more details.

• Your references may be in any style, provided that you use the consistent formatting throughout. It is essential to include author(s) name(s), journal or book title, article or chapter title (where required), year of publication, volume and issue (where appropriate) and pagination. DOI numbers (Digital Object Identifier) are not mandatory but highly encouraged. The bibliography software package EndNote, Zotero, Mendeley, Reference Manager are recommended.
• When your manuscript reaches the revision stage, you will be requested to format the manuscript according to the journal guidelines.

5) Research Manuscript Sections

Introduction: The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance, including specific hypotheses being tested. The current state of the research field should be reviewed carefully and key publications cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the main conclusions. Keep the introduction comprehensible to scientists working outside the topic of the paper.
Materials and Methods: They should be described with sufficient detail to allow others to replicate and build on published results. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited. Give the name and version of any software used and make clear whether computer code used is available. Include any pre-registration codes.
Results: Provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.
Discussion: Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible and limitations of the work highlighted. Future research directions may also be mentioned. This section may be combined with Results.
Conclusions: This section is not mandatory but can be added to the manuscript if the discussion is unusually long or complex.
Patents: This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.
Back Matter
Supplementary Materials: Describe any supplementary material published online alongside the manuscript (figure, tables, video, spreadsheets, etc.). Please indicate the name and title of each element as follows Figure S1: title, Table S1: title, etc.
Author Contributions: Each author is expected to have made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work; or have drafted the work or substantively revised it; AND has approved the submitted version (and version substantially edited by journal staff that involves the author's contribution to the study); AND agrees to be personally accountable for the author's own contributions and for ensuring that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and documented in the literature.
For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; Methodology, X.X.; Software, X.X.; Validation, X.X., Y.Y. and Z.Z.; Formal Analysis, X.X.; Investigation, X.X.;

Resources, X.X.; Data Curation, X.X.; Writing – Original Draft Preparation, X.X.; Writing – Review & Editing, X.X.; Visualization, X.X.; Supervision, X.X.; Project Administration, X.X.; Funding Acquisition, Y.Y

Funding: All sources of funding of the study should be disclosed. Clearly indicate grants that you have received in support of your research work and if you received funds to cover publication costs. Note that some funders will not refund article processing charges (APC) if the funder and grant number are not clearly and correctly identified in the paper. Funding information can be entered separately into the submission system by the authors during submission of their manuscript. Such funding information, if available, will be deposited to FundRef if the manuscript is finally published.

Please add: "This research received no external funding" or "This research was funded by [name of funder] grant number [xxx]" and "The APC was funded by [XXX]" in this section. Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

Institutional Review Board Statement: In this section, please add the Institutional Review Board Statement and approval number for studies involving humans or animals. Please note that the Editorial Office might ask you for further information. Please add "The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of NAME OF INSTITUTE (protocol code XXX and date of approval)." OR "Ethical review and approval were waived for this study, due to REASON (please provide a detailed justification)." OR "Not applicable" for studies not involving humans or animals. You might also choose to exclude this statement if the study did not involve humans or animals.

Informed Consent Statement: Any research article describing a study involving humans should contain this statement. Please add "Informed consent was obtained from all subjects involved in the study." OR "Patient consent was waived due to REASON (please provide a detailed justification)." OR "Not applicable." for studies not involving humans. You might also choose to exclude this statement if the study did not involve humans.

Written informed consent for publication must be obtained from participating patients who can be identified (including by the patients themselves). Please state "Written informed consent has been obtained from the patient(s) to publish this paper" if applicable.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies". You might choose to exclude this statement if the study did not report any data.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: Authors must identify and declare any personal circumstances or interest that may be perceived as influencing the representation or interpretation of reported research results. If there is no conflict of interest, please state "The authors declare no conflict of interest." Any role of the funding sponsors in the choice of research project; design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section. Viruses does not publish studies funded partially or fully by the tobacco industry. Any projects funded by industry must pay special attention to the full declaration of funder involvement. If there is no

role, please state “The sponsors had no role in the design, execution, interpretation, or writing of the study”. For more details please see Conflict of Interest.

References: References must be numbered in order of appearance in the text (including table captions and figure legends) and listed individually at the end of the manuscript. We recommend preparing the references with a bibliography software package, such as EndNote, ReferenceManager or Zotero to avoid typing mistakes and duplicated references. We encourage citations to data, computer code and other citable research material. If available online, you may use reference style 9. below.

Citations and References in Supplementary files are permitted provided that they also appear in the main text and in the reference list.

In the text, reference numbers should be placed in square brackets [ ], and placed before the punctuation; for example [1], [1–3] or [1,3]. For embedded citations in the text with pagination, use both parentheses and brackets to indicate the reference number and page numbers; for example [5] (p. 10). or [6] (pp. 101–105).

References should be described as follows, depending on the type of work:

Journal Articles:

1. Author 1, A.B.; Author 2, C.D. Title of the article. Abbreviated Journal Name Year, Volume, page range.

Books and Book Chapters:

2. Author 1, A.; Author 2, B. Book Title, 3rd ed.; Publisher: Publisher Location, Country, Year; pp. 154–196.

3. Author 1, A.; Author 2, B. Title of the chapter. In Book Title, 2nd ed.; Editor 1, A., Editor 2, B., Eds.; Publisher: Publisher Location, Country, Year; Volume 3, pp. 154–196.

Unpublished materials intended for publication:

4. Author 1, A.B.; Author 2, C. Title of Unpublished Work (optional). Correspondence Affiliation, City, State, Country. year, status (manuscript in preparation; to be submitted).

5. Author 1, A.B.; Author 2, C. Title of Unpublished Work. Abbreviated Journal Name year, phrase indicating stage of publication (submitted; accepted; in press).

Unpublished materials not intended for publication:

6. Author 1, A.B. (Affiliation, City, State, Country); Author 2, C. (Affiliation, City, State, Country). Phase describing the material, year. (phase: Personal communication; Private communication; Unpublished work; etc.)

Conference Proceedings:

7. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In Title of the Collected Work (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).

Thesis:

8. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.

Websites:

9. Title of Site. Available online: URL (accessed on Day Month Year).

10. Title of Site. URL (archived on Day Month Year).

Preparing Figures, Schemes and Tables

Viruses can publish multimedia files in articles or as supplementary materials. Please contact the editorial office for further information.

All Figures, Schemes and Tables should be inserted into the main text close to their first citation and must be numbered following their number of appearance (Figure 1, Scheme 1, Figure 2, Scheme 2, Table 1, etc.).

All Figures, Schemes and Tables should have a short explanatory title and caption.

All table columns should have an explanatory heading. To facilitate the copy-editing of larger tables, smaller fonts may be used, but no less than 8 pt. in size. Authors should use the Table option of Microsoft Word to create tables.

Authors are encouraged to prepare figures and schemes in color (RGB at 8-bit per channel). There is no additional cost for publishing full color graphics.

Images of cells and Western blots should be large enough and of sufficient resolution to see the relevant features. In addition, the uncropped, untouched, full-sized original images of Western blots should be uploaded independently in the “Original Images for Blots/Gels (ZIP/PDF)” field during submission.