



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE CIÊNCIAS DA SAÚDE  
DEPARTAMENTO DE BIOQUÍMICA E BIOFÍSICA  
PROGRAMA MULTICÊNTRICO DE PÓS-GRADUAÇÃO EM  
BIOQUÍMICA E BIOLOGIA MOLECULAR**



**RODRIGO CUNHA OLIVEIRA**

**CARACTERIZAÇÃO MOLECULAR DOS GENÓTIPOS C, F1 E BC DO VÍRUS DA  
IMUNODEFICIÊNCIA HUMANA TIPO 1 (HIV-1) CIRCULANTES NA REGIÃO  
NORDESTE DO BRASIL - ANÁLISE DA DISPERSÃO E HISTÓRIA EVOLUTIVA.**

**SALVADOR**

**2023**

**RODRIGO CUNHA OLIVEIRA**

**CARACTERIZAÇÃO MOLECULAR DOS GENÓTIPOS C, F1 E BC DO VÍRUS DA  
IMUNODEFICIÊNCIA HUMANA TIPO 1 (HIV-1) CIRCULANTES NA REGIÃO  
NORDESTE DO BRASIL - ANÁLISE DA DISPERSÃO E HISTÓRIA EVOLUTIVA.**

Tese de doutorado apresentada ao Programa Multicêntrico de Pós-Graduação em Bioquímica e Biologia Molecular – UFBA/SBBq, como requisito parcial para a obtenção do título de doutor em Bioquímica e Biologia Molecular pela Universidade Federal da Bahia.

Orientadora: Profa. Dra. Joana Paixão Monteiro  
Coorientador: Prof. Dr. Helder Takashi Imoto Nakaya

**SALVADOR**

**2023**

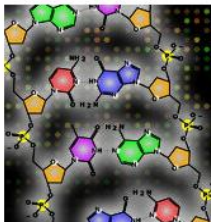
Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),  
com os dados fornecidos pelo(a) autor(a).

O48 Oliveira, Rodrigo Cunha  
Caracterização molecular dos genótipos C, F1 e BC do vírus da  
imunodeficiência humana tipo 1 (HIV-1) circulantes na região nordeste do  
Brasil – análise da dispersão e história evolutiva/Rodrigo Cunha Oliveira. –  
Salvador, 2023.  
180 f.: il.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Joana Paixão Monteiro Cunha; Coorientador: Prof.  
Dr. Helder Takashi Imoto Nakaya.  
Tese (Doutorado) – Universidade Federal da Bahia, Instituto de Ciências  
da Saúde/Programa Multicêntrico de Pós-Graduação em Bioquímica e  
Biologia Molecular, 2023.  
Inclui referências e apêndices.

1. HIV-1. 2. AIDS. 3. Subtipo C. 4. Subsubtipo F1. 5. Recombinante BC.  
6. Evolução viral. I. Cunha, Joana Paixão Monteiro. II. Nakaya, Helder  
Takashi Imoto. III. Universidades Federal da Bahia. IV. Título.

CDU 616.98(812/814)



Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq

Universidade Federal da Bahia – UFBA

Instituto de Ciências da Saúde – ICS

**PMBqBM**

Programa Multicêntrico de  
Pós-Graduação em Bioquímica e Biologia Molecular



## TERMO DE APROVAÇÃO DA DEFESA DE TESE

**RODRIGO CUNHA OLIVEIRA**

**Caracterização molecular dos genótipos C, F1 e BC do Vírus da  
Imunodeficiência Humana tipo 1 (HIV-1) circulantes na região Nordeste do  
Brasil - análise da dispersão e história evolutiva.**

**Salvador, Bahia, 12 de junho de 2023.**

COMISSÃO EXAMINADORA:

Prof. Dra. Joana Paixão Monteiro (Orientadora/UFBA)  
Presidente

Prof. Dr. Victor Figueiredo Pimentel  
Universidade Nova de Lisboa

Prof. Dra. Simone Kashima Haddad  
Hemocentro de Ribeirão Preto, Universidade Estadual de São Paulo (USP)

Prof. Dra. Fabiana Avila Carneiro  
Universidade Federal do Rio de Janeiro (UFRJ)

Márcia Barbosa da Silva

Prof. Adjunta Classe I

Departamento de Bioquímica e Biofísica  
Instituto de Ciências da Saúde - UFBA

## AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar aos meus pais Cassimiro Filho Amaral de Oliveira (*in memoriam*) e Rainice da Silva Cunha Oliveira, que sempre me apoiaram durante toda a vida, e em especial neste período, com a minha educação e formação profissional. Aos meus irmãos Marta, Marine e Júnior, pela força, apoio e incentivo pela continuidade dos meus estudos na pós-graduação.

Um agradecimento especial a professora Joana Paixão Monteiro, por me proporcionar a oportunidade de crescimento profissional quando me aceitou para orientação. Trabalhar com ela certamente me tornou o mais próximo de um “ser cientista”. Com ela pude experimentar uma nova área de estudo, que aprendi com bastante prazer. Obrigado pela parceria, paciência, dedicação, disposição no ensinar e dialogar sempre.

Aos colaboradores deste trabalho, na Fundação Osvaldo Cruz, Bahia, Tiago Gräf; no Hospital Universitário Professor Edgard Santos, Carlos Brites e na Fundação Osvaldo Cruz, Rio de Janeiro, à Luiz Alcântara, Monick Lindermayer e Marta Giovanetti e, por fim, à Filipe Ferreira de Almeida Rego, da Escola Bahiana de Medicina e Saúde Pública. Obrigado pela parceria.

Meus sinceros agradecimentos a Igor Cunha, que me acompanhou desde os estudos para a seleção do mestrado, no doutorado, pelo conforto nos momentos difíceis, me encorajando e transmitindo apoio e força nessa caminhada até agora. Muito obrigado.

Aos meus colegas na pesquisa, e em especial à Gabriela Porto, pois, com certeza, sem ela essa caminhada teria sido muito mais difícil. Obrigado por estar comigo até aqui. Agradeço ainda Juliana Sacramento e Allan Botura pela parceria, oportunidade de conviver, discutir e aprender juntos no âmbito profissional.

Aos meus professores durante minha trajetória na pós-graduação, em especial à Luzimar Gonzaga e Simone Macambira por todo o acolhimento, ensinamentos, pelos conselhos e conversas e todo o apoio me fez continuar e me proporcionou crescer profissionalmente. Obrigado.

À Dra. Isa Lyra, pelo apoio, incentivo e o encorajamento durante esses quatro anos. Muito obrigado.

Agradeço ainda aos meus amigos que foram extremamente pacientes e empáticos com a causa neste período da vida, além de o suporte prestado frente aos desafios encontrados nessa jornada.

E por fim, à Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) pela bolsa de estudos que possibilitou a continuidade dos meus estudos, para a realização desta pesquisa.

## RESUMO

**INTRODUÇÃO:** O Vírus da Imunodeficiência Humana (HIV) é o agente etiológico da Síndrome da Imunodeficiência Adquirida (AIDS), doença ainda sem vacina e sem cura. O HIV-1 apresenta uma extensa variabilidade genética e no Brasil, os subtipos B, C e o subsubtipo F1 são os genótipos prevalentes nas infecções. São detectados também recombinantes intersubtipos como BC e BF1. Os subtipos não-B tem aumentando sua frequência no território brasileiro. **OBJETIVO:** Identificar isolados virais não-B do HIV-1 (subtipo C, subsubtipo F1 e recombinante BC) na Bahia e analisar a rota de dispersão e características genotípicas e moleculares. **MATERIAL E MÉTODOS:** Foi realizada a coleta de sangue total de pessoas vivendo com HIV/AIDS acompanhadas pelo Hospital Universitário Professor Edgard Santos. O DNA genômico foi extraído e sequenciado pelo método de Sanger. Os genomas foram montados no programa computacional GENEIOUS. Alinhamentos de referência foram gerados a partir de sequências disponíveis no banco de dados Los Alamos e foram analisados em programas de Bioinformática. **RESULTADOS:** A partir da coleta de amostras foram obtidas 5 sequências do subtipo C, 19 sequências BF1 e F1 e 1 recombinante BC. As análises de filodinâmica estimam que o subsubtipo F1 alcançou o território brasileiro por volta da década de 70. O subtipo C alcançou a região Nordeste do país por volta de 1985, a partir das regiões Sul, Sudeste e Centro-oeste e os isolados do subtipo C encontrados na Bahia, descendem da linhagem encontrada na região Sul. A amostra BC de nossa coorte apresentou ponto de recombinação comum a de outras recombinantes brasileiras. A caracterização dos genomas BC do mundo revelou pontos preferencias de recombinação, com fragmento B conservado nos genes *pol*, *env* e *nef*. As análises de caracterização molecular das sequências quase completas do subtipo C, local e mundial, sugere que a maioria das cepas apresentem uma baixa capacidade replicativa. **CONCLUSOES:** Os isolados do subsubtipo F1 que circulam no Brasil pertencem a uma linhagem única, que descende do centro-oeste da África. O subtipo C encontrado na região Nordeste e estado da Bahia descende da região Sul e circula na região Nordeste desde 1985. O novo padrão da sequência BC pode representar uma nova CRF. Os resultados das análises das cepas BC revelam que características do subtipo B são mantidas nas progenies BC, podendo favorecer vantagens adaptativas. Os resultados das análises do subtipo C puro indicam que as características moleculares apresentadas parecem favorecer sua transmissão.

Palavras-chave: HIV-1, AIDS, subtipo C, subsubtipo F1, recombinante BC, evolução viral.



## ABSTRACT

**INTRODUCTION:** The Human Immunodeficiency Virus (HIV) is the etiological agent of the Acquired Immunodeficiency Syndrome (AIDS), a without a vaccine or cure. HIV-1 presents an extensive genetic variability and in Brazil, subtypes B, C and subsubtype F1 are the prevalent genotypes and intersubtype recombinants such as BC and BF1 are also detected. Non-B subtypes have increased in frequency in Brazil. **OBJECTIVE:** To identify non-B viral isolates of HIV-1 (subtype C, subsubtype F1 and BC recombinant) in Bahia and to analyze the dispersion route and genotypic and molecular characteristics. **MATERIAL AND METHODS:** Whole blood samples was collected from patients with HIV-1 followed up at the Professor Edgard Santos University Hospital. The genomic DNA was extracted and sequenced by the Sanger method. The genomes were assembled using the GENEIOUS software. Reference alignments were generated for molecular and phylogenetic analyzes from sequences available in the Los Alamos database and were analyzed through Bioinformatic softwares. **RESULTS:** Through the obtaining whole blood samples 5 subtype C sequences, 19 BF1 and F1 sequences and 1 BC recombinant sequences were obtained. The phylodynamic analyzes estimated that the F1 sub-subtype reached the Brazilian territory around the 70's. The C subtype reached the Northeast region around 1985, from the South, Southeast and Central-west regions, and the C strain found in Bahia descends from the lineage of the South region. The BC sample from our cohort showed a common recombination point to other Brazilian recombinants. The genetic characterization of BC genomes reveals a preferential recombination point with the B fragment conserved in *pol*, *env* and *nef* genes. The molecular characterization of near-full length genome of local and global subtype C suggests that the most of strain has a low replicative capacity. **CONCLUSIONS:** The F1 Brazilian isolates belongs to a single lineage, which descends from the Central-West of the Africa. The subtype C found in the Northeast region and state of Bahia descends from the South region, circulating in the region since 1985. The new pattern of BC sequence may represent a new CRF in Brazil. The genomic analyzes results of recombinant BC strains reveal that characteristics of subtype B maintained in BC strains may favor adaptative advantages. Concerning the pure C genotype, the amino acid and molecular characteristics pointing to a favor this subtype transmission.

Keywords: HIV-1, AIDS, subtype C, sub-subtype F1, BC recombinant, viral evolution.

## LISTA DE FIGURAS

<b>Figura 1</b> - Saltos interespecies – formação das linhagens do HIV.	21
<b>Figura 2</b> - Árvore filogenética do estudo de Gao e colaboradores (1999).	22
<b>Figura 3</b> - Representação da diversidade genética do HIV-1.	23
<b>Figura 4</b> - Distribuição geográfica dos subtipos, CRFs e URFs do HIV-1.	24
<b>Figura 5</b> - Estrutura e organização do genoma do HIV-1.	27
<b>Figura 6</b> - Vírião do HIV ou o vírus fora de uma célula infectada.	28
<b>Figura 7</b> - Representação do ciclo de replicação do HIV	29
<b>Figura 8</b> - Localização geográfica da Romênia.	32
<b>Figura 9</b> - Localização geográfica de Mbiju-Mayi, capital da província de Kasai Oriental, República Democrática do Congo.	34
<b>Figura 10</b> - Classificação da sequência JN692442 no banco de dados LANL.	39
<b>Figura 11</b> - Ciclo de replicação do HIV e os alvos dos antiretrovirais.	40
<b>Figura 12</b> - Fluxograma de recrutamento, coleta de amostras das pessoas vivendo HIV/AIDS acompanhados pelo Hospital Universitário Professor Edgard Santos e sequenciamento.	47
<b>Figura 13</b> - Fluxograma das análises com as sequências de nucleotídeos obtidas no Hospital Universitário Professor Edgard Santos e no banco de dados do Laboratório Nacional Los Alamos	50
<b>Capítulo I:</b>	
<i>Figure 1: Time scale Bayesian MCC tree based on pol genomic region showing the phylodynamic relationships of 671 HIV-1 subtype F1 viral isolates from worldwide.</i>	59
<i>Figure 2: Time scale Bayesian MCC tree (n=53) based on Near-Full Length Genome sequences showing the phylodynamic relationships among subtype F1 viruses circulating in Brazil (n=25), Europe (n=24) and Africa (n=3).</i>	59
<i>Figure 3: Bayesian coalescent inference of population size of subtype F1 based on 53 NFLG sequences from Brazil and worldwide.</i>	60

## Capítulo II:

**Figure 1** - Distribution map of analyzed sequences by Brazilian regions and states (N=460). 65

**Figure 2** - ML analysis of HIV-1 subtype C pol sequences (876bp) from Brazil (n=460) and worldwide (n=170). 67

**Figure 3** - Time scale Bayesian MCC tree showing phylogenetic relationships among subtype C viruses circulating in Brazil (n=455). 68

**Figure 4** - Viral dispersion dynamics of HIV-1 C in Brazil. 69

### Capítulo III:

**Figure 1** - Subtype classification of the samples 007, 011, 112, 136 and 224. 81

**Figure 2** - Maximum likelihood analyses showing phylogenetic relationships of HIV-1 subtype C sequences from Bahia with sequences from Brazil and worldwide 82

**Supplementary material 2**- Maximum likelihood analyses showing phylogenetic relationships of the HIV-1 subtype C NFLG sequence from Bahia with subtype C sequences from Brazil (25) and worldwide (1980). 84

### Capítulo IV:

**Figure 1** - Recombination patterns of HIV-1 BC strains. 91

**Figure 2** - Maximum likelihood analyses showing phylogenetic relationships (A, B, C and D) of 180 HIV-1 sequences plus four recombination BC sequences from Bahia (223) and Brazil (GQ365652, KT42778 and KT427790). 92

**Supplementary material 1**- Recombination pattern detected of the integrase fragment from the 223 sample by RDP5 program through RDP (A) and Bootscan (B) methods. 94

**Supplementary material 3**- RDP5 maximum likelihood analyses showing phylogenetic relationships (A, B, C and D) of 180 HIV-1 sequences plus four recombination BC sequences from Bahia (223) and Brazil (GQ365652, KT42778 and KT427790). 99

**Supplementary material 4**- Bootscan analysis (SimPlot) of the integrase fragment from the sample 223 against (A) KT427778, (B) KT427790 and (C) GQ365652. Sliding window of 400 base pairs moving in steps of 40 nucleotides. 102

### Capítulo V:

**Figure 1 A** - Global distribution of available BC recombinants in Los

<i>Alamos database. B - Frequencies found between Circulating Recombinant Forms and Unique Recombinant Forms.</i>	<b>118</b>
<b>Figure 2 - Recombination patterns of the 14 CRF<sub>BC</sub> circulating worldwide.</b>	<b>119</b>
<b>Figure 3 A - Preferential points of recombination in the genome of circulating and unique recombinant forms of BC HIV-1. B - Specific nucleotide positions between genes (pol, env and nef) with frequency of recombination events (<math>\geq 20\%</math>) between blue lines.</b>	<b>120</b>
<b>Figure 4- Maximum likelihood analysis of HIV-1 subtype BC sequences (7008bp) rooted by the midpoint.</b>	<b>121</b>

## LISTA DE GRÁFICOS

<b>Gráfico 1</b> - Número de casos de HIV no Brasil por ano (2017 - 2021).	<b>41</b>
<b>Gráfico 2</b> - Número (%) de casos de HIV por região brasileira.	<b>42</b>
<b>Capítulo VI:</b>	
<b>Gráfico 1</b> – Amostragem em números absolutos das 960 sequências nucleotídicas do subtipo C do HIV-1 por país.	<b>131</b>
<b>Gráfico 2</b> – Diversidade de tetrapeptídeos na alça V3 da proteína gp120 das amostras com tropismo CXCR4.	<b>132</b>
<b>Gráfico 3</b> – Frequência relativa de substituições no dipeptídeo C-S encontrados entre as 960 amostras do subtipo C do Brasil e do mundo.	<b>133</b>
<b>Gráfico 4</b> - Frequência absoluta das variações no motivo GGXXS da proteína NEF de 339 amostras do conjunto de dados.	<b>136</b>
<b>Gráfico 5</b> – Frequência relativa das substituições de aminoácidos na posição 67X em 39 (100%) cepas do subtipo C.	<b>137</b>

## LISTA DE TABELAS

### Capítulo I:

**Table 1.** *Countries of sample collection and the collected number of sequences for each dataset. CD: Democratic Republic of Congo. CF: Democratic Republic of Congo.* **57**

### Capítulo II:

**Table – 1** *Subtype C data in the Northeast Region based on phylodynamic tree* **69**

### Capítulo III:

**Table – 1** *- Socio-demographic characteristics data from HIV-1 subtype C individuals.* **80**

### Capítulo IV:

**Supplementary material 2** *- Nucleotide sequence list* **95**

### Capítulo V:

**Supplementary material 1** *– Nucleotide sequence list* **123**

### Capítulo VI:

**Tabela 1** *– Descrição da informação biológica das 26 sequências nas posições 483 e 484 na proteína GAG.* **138**

## LISTA DE APÊNDICES

<b>APÊNDICE A</b> – termo de consentimento livre-esclarecido	<b>161</b>
<b>APÊNDICE B</b> – Questionário	<b>163</b>
<b>APÊNDICE C</b> - <i>Recombination patterns of HIV-1 BC samples classified as URFs used in frequency data of study: GENOMIC CHARACTERIZATION OF HIV-1 BC RECOMBINANT VIRUSES. Oliveira, R. C. and Monteiro-Cunha, J. P. (2023)</i>	<b>164</b>

## LISTA DE SIGLAS E ABREVIATURAS

AIDS	<i>Acquired Immunodeficiency Syndrome</i> (Síndrome da Imunodeficiência Adquirida).
aLRT	<i>Approximated Likelihood Ratio Test (Approximated Likelihood Ratio Test)</i> .
ART	<i>Antiretroviral Therapy</i> (Terapia antirretroviral).
AMP	<i>Adenosine Monophosphate</i> (Adenosina Monofosfato).
BLAST	<i>Basic Local Alignment Search Tool</i> (Ferramenta Básica de Pesquisa de Alinhamento Local).
CAAE	Número da Comissão de Apresentação de Apreciação Ética .
CCR5	Receptor de quimiocina tipo 5.
cDNA	Fita de DNA complementar.
CD4 <sup>+</sup>	<i>Cluster of differentiation 4 positive</i> (Grupo de Diferenciação-4 positivo).
CEP	Comitê de Ética e Pesquisa
CXCR4	Receptor de quimiocina tipo 4.
CRF	Circulating Recombinant Form (Forma Recombinante Circulante).
dN	Taxa de substituições não-sinônimas
DNA	<i>Deoxyribonucleic acid</i> (Ácido Desoxiribonucleico).
dNTP	Desoxinucleotídeo trifosfato.
dS	Taxa de substituições sinônimas
EDTA	<i>Ethylene Diamino Tetraacetic Acid</i> (Ácido Etileno Diamino Tetracético).
ESS	<i>Effective Sample Size</i> (Tamanho Efetivo da Amostra.).
FT	Fator(es) de Transcrição
G	<i>Gamma</i> (Gama).
GMP	<i>Guanosine Mophosphate</i> (Guanosina Monofosfato).
Gp-120	Glicoproteína 120 do envelope viral.
Gp-41	Glicoproteína 41 do envelope viral.
GTR	<i>General Time Reversible</i> (Tempo Geral Reversível).
HIV	<i>Human immunodeficiency Virus</i> (Vírus da Imunodeficiência Humana).
HIV-1	<i>Human immunodeficiency Virus type 1</i> (Vírus da Imunodeficiência Humana tipo 1).
HIV-2	<i>Human immunodeficiency Virus type 2</i> (Vírus da Imunodeficiência Humana tipo 2).



HIV-1A	<i>Human immunodeficiency Virus type 1, A subtype</i> (Subtipo A do HIV-1).
HIV-1C	<i>Human immunodeficiency Virus type 1, C subtype</i> (Subtipo C do HIV-1).
HSH	Homens que fazem sexo com outros homens
HXB2	Cepa de referência do HIV-1, subtipo B.
I	<i>Invariable sites</i> (sítios invariáveis).
IN	Integrase.
INTR	<i>Nucleoside Reverse Transcriptase Inhibitors</i> (Inibidores da Transcriptase Reversa nucleosídicos).
INNTR	<i>Non-Nucleoside Reverse Rranscriptase Inhibitors</i> (Inibidores da Transcriptase Reversa não-nucleosídicos).
JpHMM	<i>Jumping Profile Hidden Markov Model</i> .
Kb	Quilobases.
LAPI	Laboratório de Pesquisa em Infectologia.
LANL	<i>Los Alamos National Laboratory</i> (Laboratório Nacional de Los Alamos).
LTR	<i>Long Terminal Repeat</i> (Repetição Terminal Longa).
ML	<i>Maximum Likelihood</i> (Probabilidade Máxima).
mM	milimolar.
MCMC	<i>Markov Chain Monte Carlo</i> (Cadeia de Markov Monte Carlo).
MCP-1	<i>Monocytes Chemotaxis Protein-1</i> (Proteína de Quimiotaxia de Monócitos-1)
NCBI	<i>National Center for Biotechnology Information</i> (Centro Nacional de Informações sobre Biotecnologia).
NF-κB	<i>Kappa-B Nuclear Transcription Factor</i> (Fator de Transcrição Nuclear <i>Kappa</i> – B).
Pmol	Picomolar.
PP	<i>Posterior Probability</i> (Probabilidade posterior).
Pr	Protease.
PreP	Profilaxia pré-exposição
RNA	<i>Ribonucleic Acid</i> (Ácido Ribonucleico).
SIV	<i>Simian Immunodeficiency Virus</i> (Vírus da Imunodeficiência Símia).
TAE	Tri-Acetato-EDTA.
<i>tMRCA</i>	<i>Time Most Recent Common Ancestor</i> (Tempo do Ancestral Comum Mais Recente).
TR	Transcriptase Reversa.

UDI	Usuários de drogas injetáveis.
UNAIDS	Programa Conjunto das Nações Unidas sobre HIV/AIDS.
µg	Micrograma.
µl	Microlitro.

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>20</b>
1.1 HIV/AIDS - Histórico e diversidade genética	20
1.2 Dados epidemiológicos	26
1.3 Estrutura genômica e ciclo de replicativo	27
1.4 Sintomatologia e aspectos clínicos da infecção	29
1.5 O subsubtipo F1, o subtipo C e o recombinante BC do HIV-1 no Brasil e no mundo	31
1.6 Repercussão da variabilidade genética do HIV-1 no tratamento e terapia antiretroviral.	39
1.7 Justificativa	41
1.7.1 Hipótese	44
<b>2 OBJETIVOS</b>	<b>45</b>
<b>3 MATERIAIS E MÉTODOS</b>	<b>46</b>
3.1 Obtenção de genomas completos de isolados virais do subtipo C do HIV-1 circulante no estado da Bahia	46
3.1.1 População e Desenho de Estudo	47
3.1.2 Coleta do sangue	47
3.1.3 Obtenção de genoma proviral	47
3.1.4 Sequenciamento	48
3.1.5 Edição dos Genomas	49
3.1.6 Determinação do subtipo das amostras	49
3.2 Busca de sequências nucleotídicas e banco de dados	49
3.3 Análises das sequências de nucleotídeos e análises e filogenéticas	50
3.4 Análise de recombinação	51
3.5 Determinação da história evolutiva e origem dos isolados do subtipo C e subsubtipo F1	52
3.6 Análise de sequências de aminoácidos	52
3.7 Considerações éticas	53
<b>4 RESULTADOS</b>	<b>54</b>

**Capítulo I:** Rastreamento a relação entre as cepas F1 do subsubtipo de HIV-1:  
uma perspectiva filodinâmica (artigo: *Tracing the relationship among HIV-1 sub-*

<i>subtype F1 strains: a phylodynamic perspective).</i>	55
<b>Capítulo II:</b> Dispersão dinâmica do subtipo C do HIV-1 em direção ao Brasil Região Nordeste. (artigo: <i>Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region</i> ).	63
<b>Capítulo III:</b> Análise molecular e filogenética do HIV-1 subtipo C no estado da Bahia, Nordeste do Brasil. (artigo: <i>Molecular and phylogenetic analysis of HIV-1 subtype C in Bahia state, northeastern Brazil</i> ).	73
<b>Capítulo IV:</b> Uma nova Forma Recombinante Circulante (CRF) do HIV-1 B/C detectada em três regiões brasileiras. (artigo: <i>A new HIV-1 B/C Circulating Recombinant Form (CRF) detected in three Brazilian regions</i> ).	85
<b>Capítulo V:</b> Caracterização genômica de vírus recombinantes HIV-1 BC (artigo: <i>Genomic characterization of HIV-1 BC recombinant Viruses</i> ).	104
<b>Capítulo VI:</b> Predição de características moleculares no subtipo C do HIV-1 encontrados no Brasil e em várias partes do mundo.	129
<b>5 DISCUSSÃO</b>	143
<b>6 CONCLUSÕES</b>	146
<b>REFERÊNCIAS</b>	149

## 1 INTRODUÇÃO

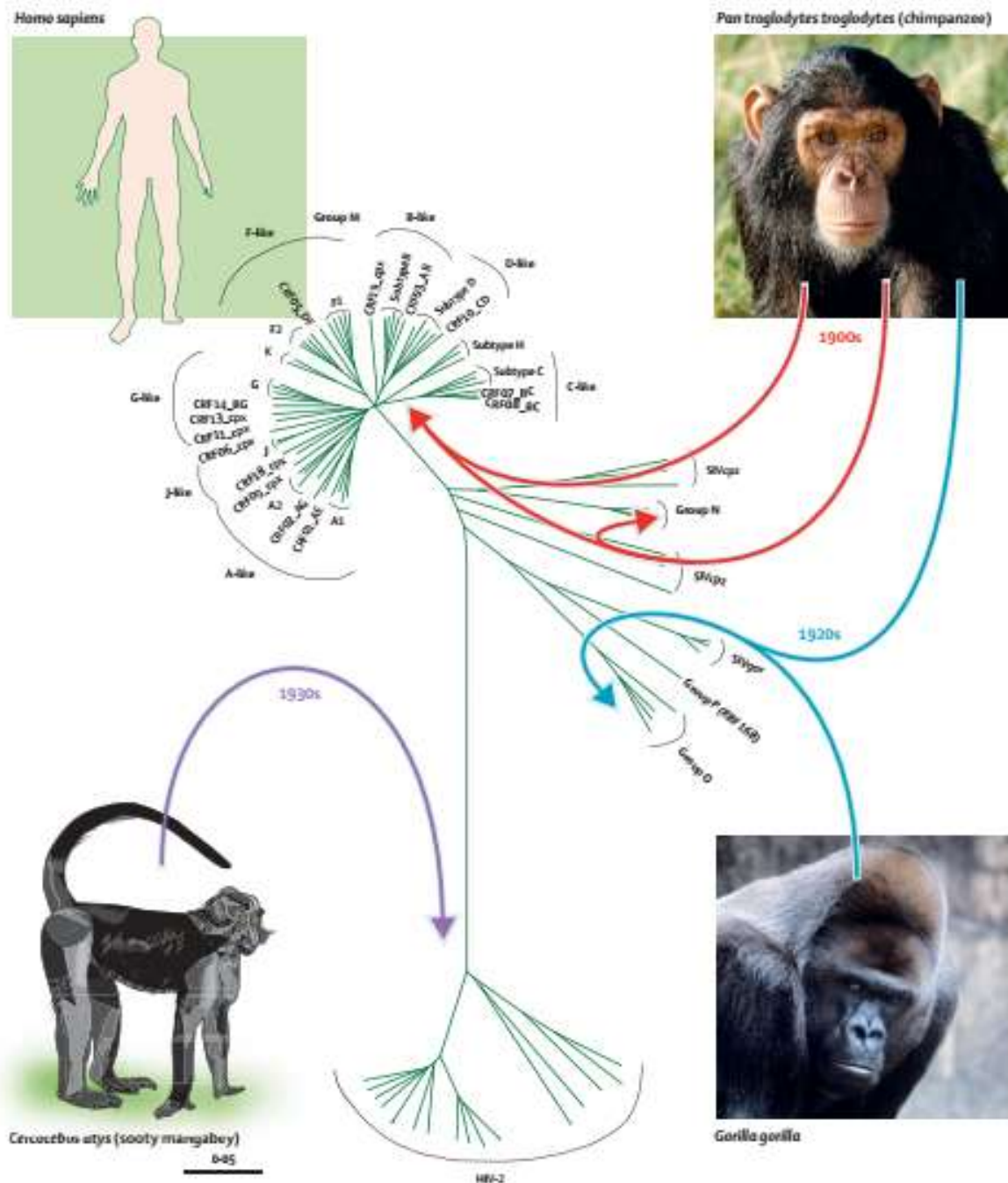
### 1.1 HIV/AIDS - Histórico e diversidade genética

O vírus responsável pela pandemia de *AIDS* (sigla em inglês para Síndrome da Imunodeficiência Adquirida), é caracterizado por conduzir uma condição de depleção da população de células do sistema imunológico, sendo estas requisitadas para infecção e sua própria replicação (IOACHIN, 1990; KATNER et al., 1987). O primeiro caso da doença foi determinado em junho de 1981 nos Estados Unidos e após seis anos, mais de 40 mil casos foram notificados no mesmo país (JAKCSON et al., 1988). Este vírus infecta, principalmente, células T CD4<sup>+</sup>, conduzindo à imunossupressão, favorecendo o aparecimento de infecções oportunistas, incluindo vírus oncogênicos, podendo cursar até ao óbito da pessoa convivendo com HIV/AIDS. Na época, os sinais iniciais relacionados a esta condição eram fraqueza, caquexia e perda de peso (KATNER et al., 1987).

Segundo a classificação biológica, o Vírus da Imunodeficiência Humana (HIV, do inglês *Human Immunodeficiency Virus*) pertence à família *Retroviridae* e ao gênero *Lentivirus*. (BOULLOSA et al., 2014). Embora hoje sejam descritos dois vírus (HIV-1 e HIV-2) (ROBERTSON et al., 2000), o HIV-1 é o responsável pelo maior número de infecções no mundo, enquanto os casos causados por HIV-2 são endêmicos, às regiões geográficas do continente africano. Este vírus (HIV-2) é ainda apresentado como imunologicamente distinto, menos comum e menos virulento (HEMELAAR et al., 2012; SANTORO et al., 2013).

Os lentivírus são detectados em mais de 30 espécies de primatas não-humanos na África sub-Sahariana (TEBLIT & ARTS, 2011). O HIV é uma zoonose, ou seja, a forma viral que tem o ser humano como hospedeiro está ligado diretamente a animais como, chimpanzés, macacos e gorilas (HEMELAAR et al., 2012). Os dois vírus (HIV-1 e HIV-2) são produtos de “saltos” (múltiplas transmissões zoonóticas) entre espécies, o HIV-1 é relacionado ao Vírus da Imunodeficiência Símia (SIV) encontrado em chimpanzés da espécie *Pan troglodytes* e gorilas da espécie *Gorilla gorilla*, enquanto o HIV-2, é descendente do vírus do macaco mangabei (*Cercocebus agilis*) (Figura 1).

Figura 1 - Saltos interespecíes – formação das linhagens do HIV.

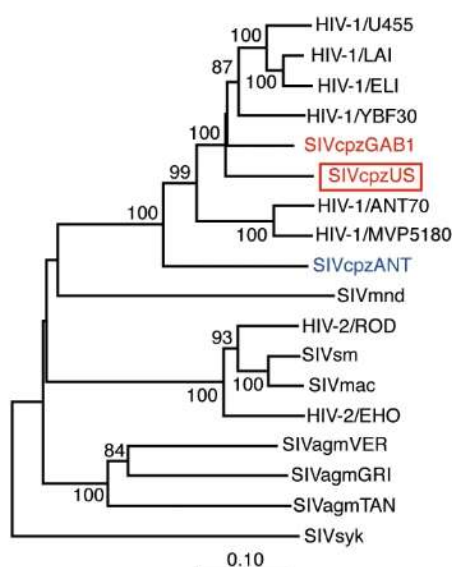


Fonte: (TEBIT & ARTS, 2011).

O estudo de Gao e colaboradores, (1999), determinou as relações de ancestralidade a partir da constatação de um grupo monofilético entre o HIV-1 e os isolados virais obtidos nos primatas em questão (Figura 2). Especula-se que a prática de caça a esses primatas, normalmente encontrados a oeste do continente africano como, na República Federativa do

Gabão e Congo, por exemplo, permitiu o primeiro contato humano com o vírus e esta transmissão provavelmente ocorreu via exposição sanguínea devido ao consumo de carne (GAO et al., 1999; SHARP et al., 2001). Estima-se que o ancestral comum ao HIV seja do início da década de 20, entre 1909 e 1930 na região de Kinshasa, República Democrática do Congo (FARIA et al., 2014), e desde então a evolução conduziu à diversificação genética, conhecida hoje pelos subtipos e subclassificações (SHARP et al., 2001; ROBERTSON et al., 2000).

Figura 2 - Árvore filogenética do estudo de Gao e colaboradores, (1999), que constatou o monofiletismo entre cepas do SIV de chipamzé e as primeiras amostras genéticas do HIV-1.



Fonte: (GAO, F. et al., 1999).

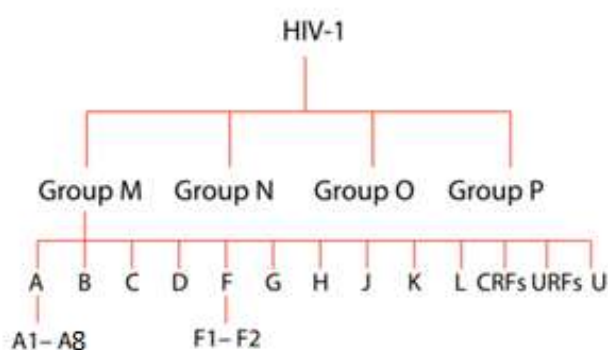
Como resultado de saltos entre espécies (*Homo sapiens sapiens x Gorilla gorilla x Pan troglodytes*) existe uma ampla heterogeneidade na classificação dos vírus pertencentes à linhagem do HIV-1, devido às variações em seu genoma, foi permitido classificar filogeneticamente os vírus deste tipo em quatro grupos: Grupo M, do inglês: *main* ou principal; O: *outlier*; N: não-M, não-O e P, cepas identificadas recentemente (CHIN, 2017; ROBERTSON et al., 2000; SANTORO et al., 2013). O grupo M do HIV-1 é responsável pela epidemia mundial, por ser mais disperso geograficamente. Sua disseminação ocorreu inicialmente entre os países do centro-oeste da África em direção ao sul e se dispersou através dos continentes (TEBIT & ARTS, 2011).

As diversas variantes do grupo M surgiram a partir dos mecanismos de evolução proporcionados pelas mutações selecionadas positivamente diante da adaptação no interior das células do hospedeiro (pressão seletiva) (KORBER et al., 2001). O mecanismo de evasão deste

vírus frente ao sistema imunológico do indivíduo, ou em resposta contra o mecanismo de ação das diversas classes das drogas antiretrovirias, conferiu-lhes alterações no código genético, polimorfismos, e isto impulsionou a diversidade de cepas cujas mutações foram selecionadas positivamente (PERELSON et al., 1998; SANTORO et al., 2013). Outro fator que favorece a evolução das diferentes formas virais está relacionado com o elevado número de replicação e a velocidade com que esse evento acontece (ROBERTS et al., 1988). A Transcriptase Reversa (TR) é a enzima responsável pela conversão da fita de RNA em dupla fita de Ácido Desoxirribonucleico (DNA). Durante a incorporação das bases nitrogenadas, a TR viral não é capaz de corrigir os erros de adição na sequência genômica produzindo cepas distintas (FABENI et al., 2017; NOMAGUCHI et al., 2017; SANTORO et al., 2013).

Essas variantes estão distribuídas ao redor do globo em regiões geográficas distintas (HEMELAAR et al., 2012; HEMELAAR et al., 2019). Esse grupo apresenta uma alta taxa de diversidade com dez subtipos catalogados (A, B, C, D, F, G, H, J, K e L), sendo o último descrito recentemente em 2019, e subsubtipos como os A1, A2, A3, A4, A5, A6, F1 e F2 (ROBERTSON et al., 2000; YAMAGUCHI et al., 2019). Em pessoas vivendo com HIV/AIDS, quando há mais de uma variante (subtipo) infectando a mesma célula, podem surgir estruturas derivadas de suas sequências genômicas, quiméricas, chamadas de Formas Recombinantes (ROBERTSON et al., 2000). Estas variantes são oriundas do processo de recombinação homóloga, baseado na caracterização de seus genomas completos, estes intersubtipos têm duas classificações: Formas Recombinantes Circulantes (CRFs) quando encontrada em pelo menos três indivíduos sem qualquer relação epidemiológica, e Formas Recombinantes Únicas (URFs), aquelas que não se encaixam no padrão descrito acima (ROBERTSON et al., 2000) (Figura 3).

Figura 3 - Representação da diversidade genética do HIV-1. “U” representa cepas que não foram subtipadas (*Untyped*).



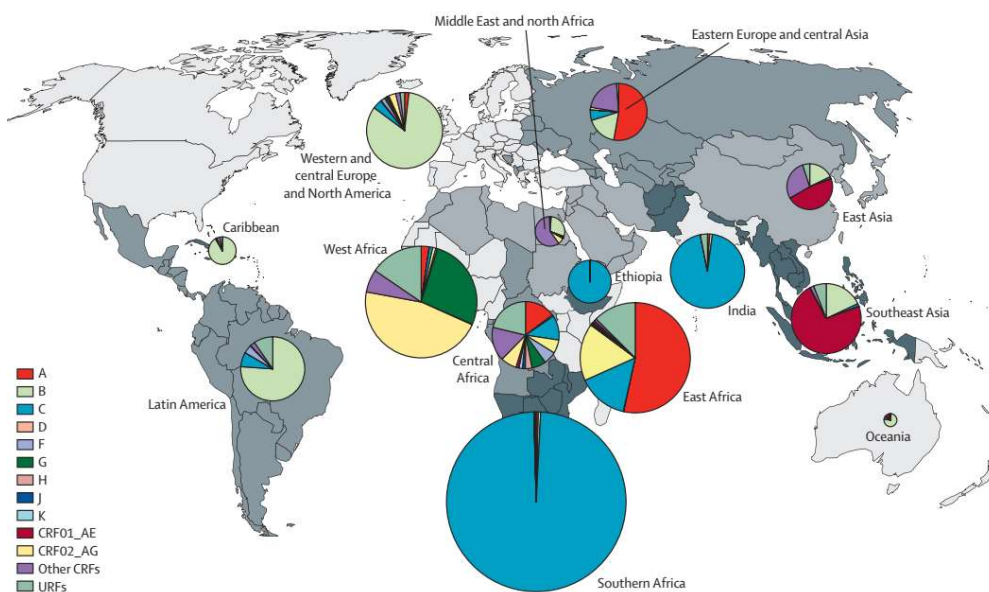
Fonte: adaptado de GIOVANETTI, M. et al., 2020.



O estudo de epidemiologia molecular de Hemelaar e colaboradores, (2019), pôde determinar o atual panorama da distribuição territorial por subtipo de HIV-1. (Figura 4). Como esperado, os países da África apresentam a maioria dos genótipos, uma vez que a HIV surgiu no continente africano, com prevalência da infecção pelo subtipo C. Esta forma também é comum em países do leste asiático. No norte da Ásia é predominantemente encontrado o subtipo A e no continente americano, europeu e oceania, o genótipo prevalente é o subtipo B. Certamente a distribuição dos outros subtipos interferem na caracterização de cada país, e essa circulação de diferentes variantes favorecem os eventos de recombinação (HEMELAAR et al., 2018).

Na base de dados do Laboratório Nacional Los Alamos (*LANL – Los Alamos National Laboratory*) são descritas 118 Formas Recombinantes Circulantes e incontáveis URFs já foram detectadas (HEMELAAR et al., 2021) Uma análise global avaliou a distribuição das cepas recombinantes (CRFs), constatando em países da África Central as formas CRF01\_AE e CRF02\_AG, África ocidental (CRF02\_AG), Brasil (CRFs\_BF1) Estônia (CRF06\_complexo), Irã (CRF35\_AD), Sudoeste asiático e China (CRF01\_AE e CRFs\_BC) (HEMELAAR et al., 2021). Enquanto as URFs predominam na Argentina, Congo, Mianmar e países da África central e ocidental.

Figura 4 - Distribuição geográfica dos subtipos, CRFs e URFs do HIV-1



Fonte: (HEMELAAR et al., 2018).

As características biológicas entre os subtipos também variam, por exemplo, a mudança

na utilização do correceptor de entrada na célula, CCR5 para CXCR4 nas fases mais avançadas da infecção por HIV-1 é um marcador para redução da população de células CD4<sup>+</sup>, critério que favorece a progressão da infecção à AIDS. Isso ocorre com a maioria dos subtipos (A, B, D, CRF01\_AE e CRF02\_AG) e essa troca é menos comum no subtipo C do vírus (HEMELAAR et al., 2012; KALU et al., 2017). Os subtipos B e D têm uma capacidade replicativa maior que o A e este, por sua vez, maior que o subtipo C. Esta característica confere aos subtipos B e D uma vantagem adaptativa (HEMELAAR et al., 2012). A via de transmissão viral também varia de acordo com a localização geográfica, considerando práticas comuns locais, por exemplo, a transmissão observada entre os Usuários de Drogas Injetáveis (UDI) e na população de Homens que fazem Sexo com Homens (HSH), é comum pelo subtipo B. Em pessoas que têm o comportamento heterossexual prevalece as formas não-B e recombinantes intersubtipos, devido, muito provavelmente às mutações selecionadas positivamente (HEMELAAR et al., 2012; HEMELAAR et al., 2013). Na África do Sul e Índia o subtipo C é prevalente, na Uganda a infecção é majoritariamente marcada pelo subtipo D e em outros países da Ásia, a CRF01\_AE é prevalente na população heterossexual (HEMELAAR et al., 2013).

De maneira geral todos os subtipos do HIV-1 apresentam sinais clínicos semelhantes nos pacientes infectados, porém a diversidade existente entre as formas genéticas pode conferir-lhes características distintas na patologia da AIDS. Estudos que avaliaram a progressão da doença em pessoas vivendo com HIV/AIDS não-B verificaram padrões clínicos distintos na infecção por diferentes subtipos e evidenciam as formas D e C como os mais agressivos, em seguida os subtipos G, AE, AG e A (SANABANI et al., 2013). O subtipo A evoluiu para um quadro onde a contagem de linfócitos T CD4<sup>+</sup> < 250 células/mm<sup>3</sup> é menos comum quando comparado aos indivíduos vivendo com HIV/AIDS do subtipo D, formas recombinantes e na infecção onde há mais de um subtipo. A progressão para a morte na condição de AIDS a partir da infecção pelo subtipo A do HIV-1 também foi menor. A expectativa de vida para as pessoas vivendo com HIV/AIDS do subtipo A foi maior em comparação aos indivíduos infectados pelo subtipo D, recombinantes e por múltiplos subtipos (KIWANUKA et al., 2010; VASAN et al., 2010). O subtipo D apresenta uma maior associação à evolução de um quadro neuropático de demência em pacientes com imunodeficiência avançada (SACKTOR et al., 2009). Este mesmo subtipo possui maior capacidade de redução das células CD4 contrastando com as contagens em indivíduos pessoas vivendo com HIV/AIDS dos subtipos B, C e CRF02\_AG, por exemplo (EASTERBROOK et al., 2010). O subtipo C foi considerado semelhante ao A quando comparadas as patogenicidades, no estudo conduzido por Vasan e colaboradores em 2006, a

progressão da doença nas pessoas vivendo com HIV/AIDS e dessas variantes não diferiram significativamente entre os dois subtipos, mas a variante D foi, novamente, mais patogênico entre os subtipos.

## 1.2 Dados epidemiológicos

Segundo os dados do UNAIDS, em quase 40 anos desde a descoberta do vírus até o ano de 2020, cerca de 37,7 milhões de pessoas no mundo vivem com HIV/AIDS, destes 28,2 milhões de pessoas possuem acesso à terapia antiretroviral (aproximadamente 74% da população). Os dados de óbitos relacionados a esta condição (HIV/AIDS) são de 36,3 milhões desde o início da epidemia e 680 mil óbitos ocorreram somente no ano de 2020. Novas infecções ao redor do mundo atingiram a marca de 1,5 milhões de casos no mesmo ano. Estratificando estes dados, aproximadamente 95,4% corresponde a população adulta (37 milhões) e 4,6% a crianças entre 0 e 14 anos. Mais da metade das pessoas que convivem com o vírus (53%) são mulheres e meninas (UNAIDS, 2021).

Segundo o Ministério de Saúde Brasileiro o primeiro diagnóstico de AIDS relatado no Brasil aconteceu na década de 80, mais precisamente no ano de 1982. Somente em 2020 foram computados mais de 32 mil novos casos de infecção por HIV e mais de 10 mil mortes foram relacionadas a condição de AIDS no mesmo ano no Brasil (BRASIL, MINISTÉRIO DA SAÚDE, 2021). De maneira geral, mais de 381 mil casos de HIV já foram diagnosticados desde que a notificação tornou-se compulsória (junho 2007 a junho 2021). No Brasil a maior parte da população com esta condição são homens (69,8% - 266.360) e 115.333 (30,2%) mulheres.

Ao longo dos anos, 141.025 mulheres gestantes foram detectadas positivas para o HIV no país, 5,5% (7.814) somente no ano de 2020. Isso revela uma tendência da infecção no país além da transmissão HSH, a heterossexualização do HIV e aumento dos casos via transmissão vertical. Os dados apontam para um incremento principalmente nas regiões Norte e Nordeste (BRASIL, MINISTÉRIO DA SAÚDE, 2021).

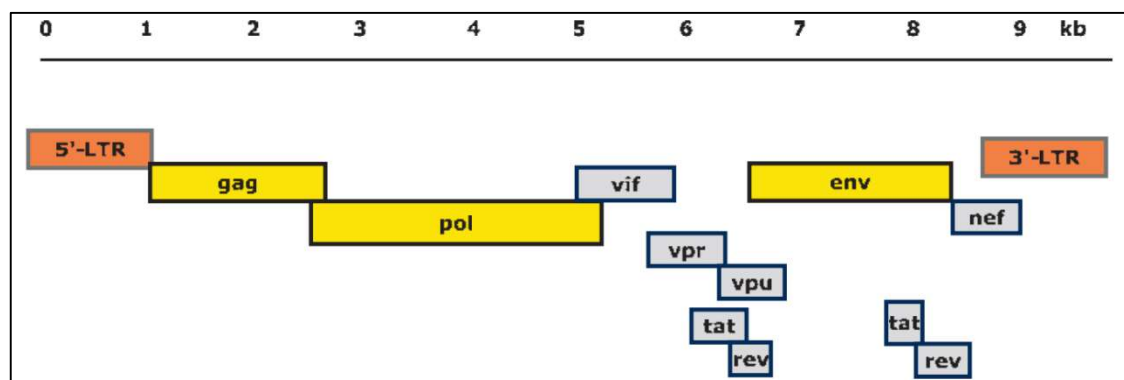
A Profilaxia pré-exposição (PrEP) foi implementada no país em 2017, isto configura um avanço na prevenção contra o HIV, principalmente em populações consideradas alvo (homossexuais, homens que fazem sexo com homens, profissionais do sexo, pessoas que usam drogas injetáveis, carcerários, casais soro-discordantes, ou de prática anal e sem preservativo em geral) (SOUSA et al., 2021). Estes medicamentos tem sido utilizados como tentativa de diminuição do risco de adquirir o vírus, e por consequência, reduzir o número de casos a partir

do bloqueio de determinadas etapas do ciclo viral (SOUSA et al., 2021).

### 1.3 Estrutura genômica e ciclo de replicativo

O genoma do Vírus da Imunodeficiência Humana é composto de duas fitas simples de ácido ribonucleico (RNA) contendo mais de 9 quilobases (kb) nucleotídicas distribuídas em três grandes genes que codificam as proteínas estruturais do envelope (*env*), do capsídeo (*gag*), e as enzimas virais, a Integrase (In), Protease (Pr) e Transcriptase Reversa (RT), (*pol*). Há ainda seis outros genes que codificam as proteínas acessórias *vif*, *vpr*, *vpu* e *nef* e as proteínas regulatórias *rev* e *tat*. (Figura 5). O genoma é delimitado em suas extremidades por duas regiões LTR (do inglês, *Long Terminal Repeat*), as quais, em conjunto com a proteína *tat*, são responsáveis pelo controle da transcrição para produção de RNA mensageiro, proporcionando a replicação a partir da integração dos genomas viral e do hospedeiro (FEINBERG et al., 1986; HARRICH et al., 1997).

Figura 5 - Estrutura e organização do genoma do HIV-1. São mostrados os quadros de leitura dos genes que codificam proteínas para estrutura e regulação: LTR = longa repetição terminal; *gag* = antígeno específico do grupo; *pol* = polimerase; *env* = envelope. No caso dos genes reguladores, as proteínas de *tat* e *rev* são compostas por duas regiões gênicas. As sequências de ácido nucleico de 5' e 3' LTR não são traduzidas em proteína. O genoma é composto por 9.200 - 9.600 nucleotídeos no caso do HIV-1.

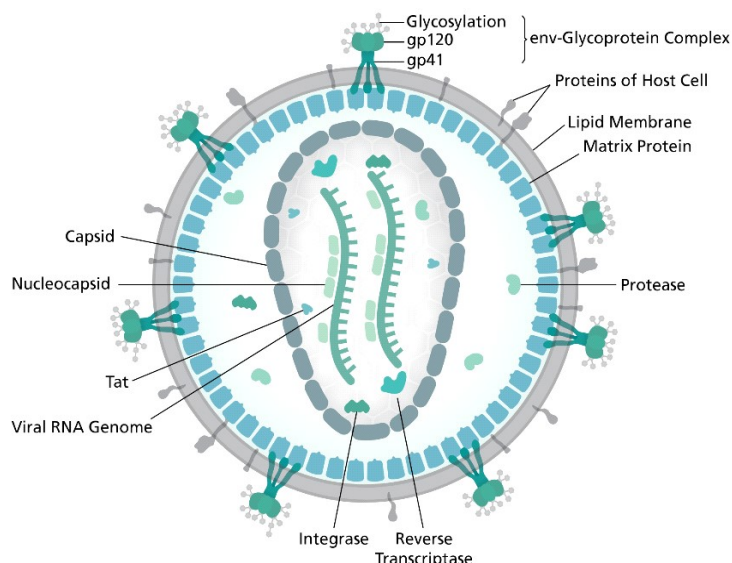


Fonte: (GÜRTLER, L. et al., 2016).

Toda essa maquinaria está envolta em um capsídeo que é revestido por bicamada lipídica. Na superfície externa da membrana são expressas proteínas que são relacionadas com o ciclo replicativo e infeccioso do vírus, as proteínas *gp120* e *gp41* como ilustrado na Figura 6. (ABBAS et al., 2015; GELDERBLOM et al., 1988). Estas duas últimas moléculas possuem alta especificidade para determinados receptores e coreceptores encontrados no linfócito T. A

proteína CD4 e os coreceptores de quimiocinas CXCR4 e CCR5, os mais comumente utilizados por este vírus, são fundamentais para o seu acesso ao interior das células (CLAPHAM et al., 2002). A infecção ocorre após a adosorção entre receptor CD4 e um dos coreceptores (CCR5 ou CXCR4) do linfócito T com o complexo proteico viral gp120-gp41 (interação proteína-proteína).

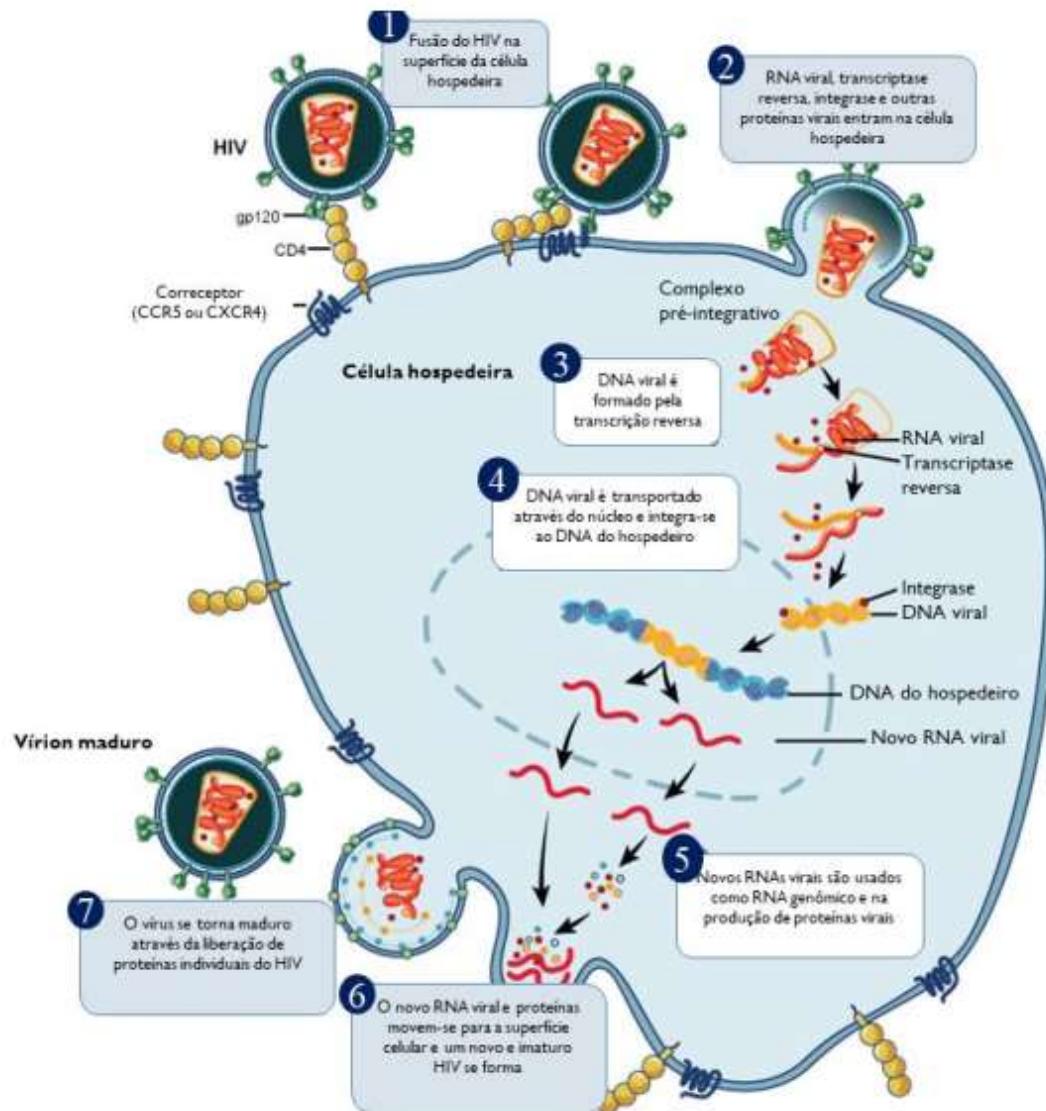
Figura 6 - Virion do HIV ou o vírus fora de uma célula infectada.



Fonte: <http://bigthink.com/philip-perry/we-may-soon-see-the-end-of-hiv>.

O domínio C4 da gp120 interage diretamente com a estrutura CD4, induzindo mudanças conformacionais na estrutura da proteína, viabilizando um sítio adicional de ligação entre a gp120 e os correceptores de quimiocinas (CCR5 ou CXCR4), então a porção N-terminal da gp41 forma um canal de conexão até a membrana da célula do hospedeiro (LUCIW et al., 1996). Em seguida acontece a fusão de membranas entre as estruturas do vírus e da célula humana, havendo a penetração (condução do capsídeo até o citoplasma). A próxima etapa ocorre após a fagocitose do capsídeo por um endossomo, proporcionando o desnudamento do material genético e das enzimas, TR, Pr e IN, dando início às suas funções. A Transcriptase Reversa age transcrevendo a fita de RNA em DNA complementar (cDNA), que é conduzida até o núcleo e a Integrase insere o DNA proviral ao genoma do hospedeiro, e a transcrição e do genoma proviral dependem, agora, do ciclo celular para ocorrer (SOUSA et al., 1993). Após as etapas de tradução das proteínas não estruturais, favorecendo o próprio ciclo de biossíntese viral, e estruturais, para que haja a montagem de novos vírions, e esta última etapa acontece a partir da ação da Protease.

Figura 7 - Representação do ciclo de replicação do HIV (tradução livre).



Fonte: <https://www.niaid.nih.gov/diseasesconditions/hiv-replication-cycle>

#### 1.4 Sintomatologia e aspectos clínicos da infecção

A exposição a fluidos biológicos de pacientes soropositivos como, sangue, sêmen, por exemplo, é a via de contágio predominante para este vírus (JACKSON et al., 1988). No início da epidemia a principal população atingida eram os portadores de hemopatias receptores de transfusão, a população HSH e UDI (IOACHIM et al., 1990). Segundo dados do Ministério de Saúde Brasileiro, (2021), as infecções por HIV nos últimos dez anos no país têm acometido a população jovem entre 20 e 34 anos, principalmente em homens que se relacionam entre si.

A patologia se caracteriza, de maneira geral, pela infecção latente, com quadro clínico

de astenia e linfadenopatias (BARRÉ-SINOUSI et al., 1983). Podem ser verificadas pelo menos três fases. A fase aguda é normalmente descrita com a presença de sintomas semelhantes aos apresentados pelas infecções por outros vírus, como a gripe, por exemplo. Febre, dores de garganta, mal-estar e *rush* cutâneo são as queixas comuns nessa etapa inicial, que é marcada também pela produção maciça de novos vírus (IAOCHIN, 1990). O estudo de Hoenigl e colaboradores, (2016), observou em 90 participantes vivendo com HIV/AIDS, diagnosticados com fase aguda, os sinais e sintomas compartilhados nesta etapa. Os autores verificaram que as manifestações clínicas freqüentes foram: febre e mialgia e menos comum: artralgia e perda de peso. Ainda pode-se observar fadiga, dor de cabeça, suor noturno, faringite, e manifestações gastrointestinais (náuseas, vômitos e diarréias) (HOENIGL et al., 2016).

A segunda fase, a crônica, é mais duradoura e apresenta sintomas variados e inespecíficos (IOACHIM et al., 1990). Neste ponto os sinais relacionados à infecção são “silenciosos”. Há produção de gamaglobulinas anti-HIV que podem neutralizar partículas virais na circulação e evidenciar um período livre de sintomas aparentes (IOACHIM et al., 1990). A terceira fase da infecção é marcada pelos sintomas desencadeados pela Síndrome da Imunodeficiência Adquirida. Esta etapa é caracterizada pela supressão do sistema imunológico e a elevação no número de vírus na circulação, sendo estes os sinais correlacionados com a progressão da doença. A partir da capacidade do HIV-1 de infectar células brancas, principalmente os linfócitos T CD4<sup>+</sup> e por conta da atividade necessária para o seu ciclo de replicação nesta célula, ocorre o efeito de imunossupressão com consequente avanço da infecção após a destruição das células CD4<sup>+</sup> e liberação de novos vírus na corrente sanguínea. A redução na contagem global dos linfócitos principalmente prejudica as respostas imunes (inata e adaptativa celular e humoral), tornando o indivíduo susceptível ao surgimento de enfermidades oportunistas como a infecção pelo Herpes Vírus Humano-8 (HHV-8), agente etiológico do Sarcoma de Kaposi, condição clínica observada mais frequente associada à condição de HIV/AIDS, principalmente no início da epidemia no início dos anos 80 (ABBAS et al., 2015; KATNER et al., 1987).

As infecções recorrentes estão associadas às principais causas de morte ao paciente com diagnóstico de AIDS. A aquisição de bactérias que colonizam o trato respiratório, como o *Micobacterium tuberculosis*, *Streptococcus pneumoniae* e *Haemophilus influenza* são comuns e vírus, a exemplo do Citomegalovírus e o *Epstein-Barr*, podem acometer quadros neoplásicos (IOACHIM, 1990). Com o advento da terapia antiretroviral, que elevou a expectativa de vida das pessoas vivendo com HIV/AIDS, houve uma mudança nas causas de morte por HIV-1

(WING, 2017). A redução da carga viral e consequente manutenção do sistema imune do hospedeiro favoreceram diminuição dos quadros de infecções e tumores nas pessoas vivendo com HIV/AIDS, aumentando a sobrevivência e permitindo alcançar idades mais avançadas, o que não era comum antes da terapia farmacológica, com isso outras causas de morte relacionada à condição de HIV/AIDS surgiram (MPONDO, 2016). Problemas cardiovasculares, hepáticos, renais e metabólicos foram diagnosticados nessa população em uso crônico dos antiretrovirais, sendo também, motivos que conduzem as pessoas vivendo com HIV/AIDS ao óbito. Estas novas condições clínicas observadas foram relacionadas como desafio do envelhecimento nas pessoas vivendo com HIV/AIDS (WING, 2017).

O HIV-1 também expressa tropismo às células do sistema nervoso desencadeando processos neuropáticos, como disfunções motoras, alterações de memória, do equilíbrio e até demência. Os processos patológicos relacionados ao tecido neural baseiam-se no efeito de desmielinização das células nervosas. Os monócitos, macrófagos e outras células infectadas pelo HIV parecem liberar quimiocinas e outras enzimas proteolíticas que destroem a bainha de mielina, contribuindo para o aparecimento destes sintomas (IOACHIM, 1990).

A elevação da carga viral no plasma do hospedeiro é efeito da imunossupressão e está diretamente relacionado com a capacidade do indivíduo em transmitir o vírus. A manutenção da contagem de linfócitos ( $>250/\text{mm}^3$ ) é crucial para o controle da infecção que pode ser feita com auxílio da terapia antirretroviral (ABBAS et al., 2015; IOACHIM, 1990; JACKSON et al., 1988).

### 1.5 O subsubtipo F1, o subtipo C e o recombinante BC do HIV-1 no Brasil e no mundo

Sabe-se que o subtipo B do HIV-1 é a variante mais comum encontrada no território brasileiro (67%), porém já são descritos o aumento de outras formas como os subtipos C (14%) e F (10%), além de recombinantes BC, BF e outros subtipos e recombinantes em menor proporção (9%) (D, CRF02\_AG, CRF45\_cpx [BCF1]) (ALCÂNTARA et al., 2013; GRAF et al., 2021; SILVEIRA et al., 2012). De acordo com o trabalho de Bello e colaboradores, (2007), o subtipo B do HIV-1 alcançou o território brasileiro entre o período que corresponde às décadas de 60 e 70 e sua origem seja resultante de múltiplas introduções. Há dados que suportam a teoria de inserção do subsubtipo F1 no Brasil 10 anos mais tarde, a partir de uma introdução única e essa forma é proveniente da região central do continente Africano (AULICINO et al., 2007; BELLO et al., 2007). O subsubtipo F1 foi descrito em 1994 na



Romênia em crianças vivendo com HIV/AIDS (DUMITRESCU et al., 1994). De acordo com os trabalhos de Bandea e colaboradores, (1994), e Metha e Colaboradores, (2011), a Romênia é o país da Europa que apresenta as maiores taxas de prevalência de F1 nas infecções por HIV-1, a Bulgária, Bélgica e Espanha aparecem logo atrás (BANDEA et al, 1994; PERNAS et al., 2014). A cidade de Kinshasa na República Democrática do Congo (DRC) (região central do continente africano) é apontada com local de origem da maioria dos genótipos do HIV-1, incluindo o F1 (FARIA et al., 2019) e estudos filogenéticos evidenciam uma estreita relação da epidemia pelo F1 encontrado na Angola, país do centro-oeste do continente africano e a Romênia, localizada na região leste do continente europeu (Figura 7). As análises filodinâmicas estimam o ano de 1984 para a detecção do ancestral comum mais recente neste país (Romênia), enquanto o estudo no território angolano é datado de 1975, sugerindo fortemente migrações entre Angola e Romênia nos dois continentes (METHA et al., 2011).

Figura 8 - Localização geográfica da Romênia (marcador em vermelho).



Fonte: Google Maps (Adaptado).

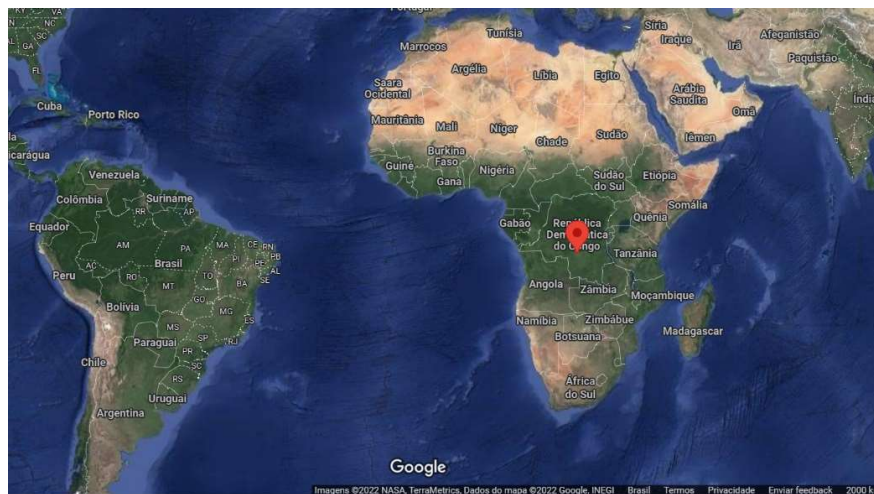
No Brasil, esta variante foi detectada por volta do fim dos anos 80 e início da década de 90 (Louwagie et al., 1994) e sua ancestralidade também é filogeneticamente vinculada com o centro-oeste da África (GUIMARÃES et al., 2009). Esta detecção está intimamente relacionada com um grande movimento histórico migratório de angolanos ao Brasil devido a uma guerra civil naquele país, e foi estimado que cerca de 15.000 pessoas tenham se abrigado no território

brasileiro entre as décadas de 70 e 90 (BAENINGER et al., 2000). O subsubtipo F1 experimentou uma rápida expansão nos 20 anos seguintes após sua chegada em território brasileiro (meados da década de 70) e de acordo com o estudo evolutivo de Bello e colaboradores, (2007), após a década de 80 apresentou uma diminuição na taxa de crescimento populacional, o que refletiu nos boletins epidemiológicos da época (BELLO et al., 2007). Dados apresentados em estudos de epidemiologia molecular sugerem que o F1 seja mais prevalente nas regiões Norte (12%) e Sudeste (12%), as demais regiões tem as taxas: Centro-Oeste (10,9%), Nordeste (10,6%) e Sul (3,3%) (ARRUDA et al., 2018). Foi especulado que a diferença de 10-15 anos entre a introdução do subtipo B e a variante F1 no país justifique as diferenças de prevalência entre estes genótipos na população (BELLO et al., 2007), embora a co-transmissão destas variantes ainda favoreçam o surgimento de formas recombinantes BF1. O Brasil possui o maior número de CRFs envolvendo ambos os subtipos já descritas, além de outros padrões de recombinações únicos identificados (SOUZA et al., 2019). Para ilustrar esta observação, um estudo de diversidade genética destes vírus identificou 1,6% do genótipo F1, enquanto formas recombinantes BF1 alcançou a taxa de 10,1% das infecções por HIV-1 em três estados da região Centro-Oeste do Brasil (Goiás, Mato Grosso e Mato Grosso do Sul) contrapondo o F1 puro (STEFANI et al., 2007).

Sanabani e colaboradores, (2006), especularam que a diminuição das taxas de F1 nas infecções no país e o aumento de padrões de recombinação entre B e F1 possa ter conferido uma ainda desconhecida vantagem seletiva. Nesta análise, os autores verificaram que nenhum dos mosaicos entre B e F1 obtidas nas novas amostras correspondiam ao padrão da, então identificada, CRF12\_BF1, sugerindo que estas tenham surgido de eventos de recombinação independentes, muito embora apresentem pontos de recombinação comuns. Contudo, foi possível concluir que os recombinantes BF1 é o segundo clado mais comum no Brasil, atrás apenas do subtipo B puro (SANABANI et al., 2006). Uma análise realizada no estado da Bahia, região Nordeste do Brasil, revelou um aumento na prevalência de F1 ao longo dos anos: 2,3% em 2002, 4,3% em 2006, 6,9% em 2007 e 14,4% em 2012; enquanto os BF1 variou 13,1% em 2002 a 24,1% em 2007 (AMARAL et al., 2017). No estado de Pernambuco, também na região Nordeste, o subsubtipo F1 é responsável por 81,5% das infecções por HIV-1 entre os genótipos não-B (LIMA et al., 2016). Em outra pesquisa, foi verificada a prevalência de 31,7% do subsubtipo F1 e a análise de expansão populacional sugere um crescimento deste genótipo nas infecções na região Nordeste, principalmente no estado de Pernambuco, onde o F1 alcançou o território por volta do ano de 1982 (LIMA et al., 2017).

Dados de pesquisa filogenética sugerem que o subtipo C surgiu por volta da década de 50 em Mbuji-Mayi, cidade da província de Kasai Oriental, na República Democrática do Congo (Figura 8) (FARIA et al., 2019). De acordo com os dados históricos, esta cidade experimentava um grande crescimento econômico, devido à exploração de minérios na região, atraindo imigrantes de outras localidades do sul do país, favorecendo a disseminação da cepa C até o leste e sul do continente (DALAI et al., 2009; GARTNER et al., 2020). O subtipo C foi isolado em um paciente na Etiópia, no ano de 1986 (AYEHUNIE et al. 1990) e sua sequência genômica completa foi publicada apenas 10 anos mais tarde (SALMINEN et al., 1996). O subtipo B é atualmente a forma mais dispersa mundialmente, porém a forma genética C é encontrada em mais da metade das infecções ao redor do globo e sua predominância geográfica está relacionada as áreas populosas como em países da região Sul do continente africano e Índia (JONES et al., 2009). Conforme cita Graf e colaboradores (2015) estudos mais recentes notificaram essa variante na Europa, Austrália, China e América do Sul.

Figura 9 - Localização geográfica de Mbuji-Mayi, capital da província de Kasai Oriental, República Democrática do Congo (marcador em vermelho).



Fonte: Google Maps (Adaptado).

A predominância do subtipo C nas infecções pode estar associada suas características biológicas, que possui menor aptidão em sua replicação em comparação a outros subtipos (BALL et al., 2005; VASAN et al., 2006). Essa perda de capacidade replicativa parece conduzir uma lenta progressão da doença em pessoas vivendo com HIV/AIDS deste subtipo, o que favorece a manutenção de células infectadas, oportunizando a transmissão por fluidos biológicos e através da mucosa genital. Deste modo, as chances de transmissão é aumentada,

uma vez que vírions livres tem meia-vida menor quando comparada as célula íntegras, que continuam liberando as partículas virais após a transmissão, refletindo em sua prevalência (KIGUOYA et al., 2017). Devido a alta taxa de erro na incorporação de bases nucleotídicas durante a replicação do genoma viral, associada a ausência de um mecanismo de correção de incorporação de nucleotídeos, as mutações, de modo aleatório, podem conduzir à seleção positiva das mutações, favorecendo adaptações entre os subtipos e ainda, entre as cepas do mesmo genótipo (BALL et al., 2005; VASAN et al., 2006). Devido a estas propriedades virais, provavelmente, a infecção pelo subtipo C é mais comumente relacionada a transmissão pela via de contato sexual, entre pessoas do sexo oposto (SANTOS et al., 2015). Isso foi associado ao forte tropismo que o subtipo C apresenta à região da mucosa vaginal, sendo esta, a principal via de infecção em recém-nascidos. Essa característica parece não ser referente aos outros subtipos como o B, por exemplo, que está relacionada a infecção entre pessoas do mesmo sexo (ALCÂNTARA et al., 2013; SILVEIRA et al., 2012).

Outra característica é a preferência da utilização do correceptor CCR5, independentemente da fase da infecção, o que parece estar relacionado com um maior período assintomático e elevadas contagens de linfócitos T CD4<sup>+</sup>. Esta propriedade parece favorecer a transmissibilidade e a difusão desse subtipo pelo mundo. (BOULLOSA et al., 2014; KALU et al., 2010). Outra particularidade do vírus desse subtipo é a existência, nas regiões LTR, de pelo menos três sítios de ligação para o fator de transcrição (FT) NF-κB, que é essencial para o início da multiplicação viral, comparativamente, isolados do subtipo B apresentam apenas dois sítios de ligação para esta molécula (HUNT et al., 2001). Já foram detectadas cepas C apresentando o quarto sítio de ligação, inclusive no Brasil (BOULLOSA et al., 2014). Para o subtipo C, estes sítios apresenta um grau de variabilidade genética, que ao contrário de outros subtipos mostram sítios altamente conservados (BHANGE et al., 2021). Apesar de ainda não haver conhecimento específico sobre os mecanismos e os efeitos biológicos dessa aquisição de sítios, sabe-se que podem estar relacionados à coinfeção e uso de drogas injetáveis (METHA et al., 2021). A regularidade com que um terceiro sítio para o fator de transcrição NF-κB se apresenta no subtipo C pode estar relacionado a condições de danos cognitivos. O tropismo marcante deste vírus pelas células do sistema nervoso central pode ser referente a potenciais sítios de ligação de fatores de transcrição encontrados nas células deste sistema (FERRARO et al., 2013). Além disso, a duplicação de FT não se limita ao NF-κB, outros fatores como, AP-1 (Proteína Acopladora-1) e RBE III (alvo do RBF-2) também são comumente duplicados e parecem estar envolvidos com ciclo de latência deste vírus (BHANGE et al., 2021). Especula-se que a

duplicação de FT seja um mecanismo compensatório aos mecanismos de virulência atenuados do subtipo C, isto é, taxa de fusão e atividade enzimática reduzida de retrotranscrição e montagem viral em relação aos outros subtipos (GARTNER et al., 2020).

A história do subtipo C no país não difere das outras variantes mais comuns. Em 2005, o estudo de Salemi e colaboradores mostrou uma introdução única do subtipo C a partir da região Sul. Em seguida, foi demonstrado que este subtipo seria derivado de países do leste da África, como Burundi e adjacências (BELLO et al., 2008; DELATORRE et al., 2013). Há relatos na literatura descrevendo que essa inserção pode ter ocorrido entre os países africanos e o Brasil diretamente, ou de forma indireta, sugerindo que as sequências virais do subtipo C, oriundas desses mesmos países, tenham rota pelo Reino Unido antes de chegar a território brasileiro (BELLO et al., 2012; OLIVEIRA et al., 2010). Análises filogeográficas mais recentes evidenciaram pelo menos cinco possíveis entradas adicionais do subtipo C no Brasil e esses eventos posteriores parecem ter ocorrido a partir da região Sudeste (DELATORRE et al., 2013). A existência desse subtipo no país foi determinada por volta dos anos 90, quando foram isoladas suas sequências em três pacientes, sendo dois destes da cidade de Porto Alegre (região Sul) e um de São Paulo (Sudeste), porém acredita-se que o subtipo C já circulava no país entre meados da década de 60-70 (ALCALDE et al., 2012; BELLO et al., 2008; DELATORRE et al., 2013). A região Sul apresenta co-circulação entre as formas B e C e cerca de 50% de infecções por HIV-1 são causadas pelo subtipo C e recombinantes BC a depender do estado (BELLO et al., 2008; GRAF et al., 2015). Após a inferência da introdução do subtipo C no território brasileiro, possivelmente no estado do Paraná, este vírus parece ter migrado para Santa Catarina e Rio Grande do Sul, por conta da proximidade geográfica, e até os estados mais ao norte do país. (BELLO et al., 2012; DELATORRE et al., 2013; VÉRAS et al., 2011).

Sendo assim, a epidemia brasileira do subtipo C teve origem na região Sul e tem se dissipado por toda extensão territorial do país, sendo este subtipo encontrado na região Centro-Oeste, por conta da rota migratória favorecida pela oferta de emprego no campo da agricultura. A pesquisa realizada por Alcântara e colaboradores em 2013, constatou que as amostras do subtipo C do estado de Goiás assemelham-se às encontradas em outros estados, principalmente as cepas encontradas nos estados do Sul e Sudeste do Brasil. A região Sudeste, por sua vez, também já foi associada com a dispersão desse subtipo para além dos estados mais ao sul do país. Nessa região, o estado de São Paulo foi relacionado como ponto secundário na rede de dispersão do vírus no país, por se tratar do principal polo urbano nacional e trajeto de viagens,

o que permite conexão geográfica com outros estados (ALCÂNTARA et al., 2013; BRÍGIDO et al., 2008; DELATORRE et al., 2013).

Essa dispersão viral pode ser verificada através de estudos publicados que notificaram o subtipo C em outros estados das regiões Norte e Nordeste. Nos trabalhos de Machado e colaboradores (2017) e Costa e colaboradores (2016) foram verificadas 3,3% e 3%, respectivamente, do subtipo C nos estados da região Norte. As sequências desse subtipo encontradas nos estados do Amazonas e Tocantins possuem relações filogenéticas robustas quando analisadas juntamente com outras amostras de estados mais ao sul. O estado do Tocantins parece ter, ainda, o papel de condutor da difusão viral nesta região. (CARVALHO et al., 2011; CUNHA et al., 2012). Em pesquisas realizadas na região Nordeste como, a publicação de Delatorre e colaboradores em 2017, já é possível observar a existência do subtipo C. Seu aparecimento foi relacionado com o aumento no número de casos de HIV nos últimos anos e algumas das causas associadas ao crescimento podem estar relacionadas com o turismo, prostituição, carência econômica e social, o que favoreceu o incremento de formas não-B, principalmente do subtipo em questão. Foi publicado um estudo realizado no estado do Ceará, no qual detectou-se a prevalência do subtipo C em 5,4% (ARRUDA et al., 2011). Outro levantamento realizado por Monteiro-Cunha e colaboradores em 2009 e 2017, constatou pessoas vivendo com HIV/AIDS pertencentes ao subtipo C, em cidades interioranas e na capital do estado da Bahia. Esses dados reforçam a hipótese da rota de dispersão do subtipo C dos estados da região Sul rumo ao Norte do Brasil (GRAF et al., 2013).

As primeiras formas recombinantes circulantes BC (CRF\_07 e CRF\_08) foram descritas no ano 2000, na China (SU et al., 2000), embora este país tenha alcançado o auge da epidemia de HIV-1 antes da década de 90, com dois principais subtipos em circulação local, o subtipo B e C (HENG et al., 1996). Das 14 CRFs BC já descritas, 11 (CRF\_07, CRF\_08, CRF\_57, CRF\_61, CRF\_62, CRF\_64, CRF\_85, CRF\_86, CRF\_88, CRF\_110 e CRF\_118) foram identificadas na China (<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). Han e colaboradores (2013), sugerem que devido a alta prevalência destes subtipos (B e C) e ao estabelecimento destas CRFs associado ao elevado número de URFs já identificadas, os eventos de recombinação de segunda geração, aqueles que envolvem recombinantes como parentais, representam um propulsor para a alta diversidade genética de cepas existentes neste país, justificando o aparecimento de novas URFs (ZHANG et al., 2019) e pelo menos duas CRFs no último ano (2021), as CRF\_110 (LI et al., 2021) e CRF\_118 (GAO et al., 2021).

Além da China, a Espanha também registrou recentemente uma nova CRF envolvendo

os dois subtipos: CRF\_108 (CAÑADA et al., 2021). Os autores relatam que a forma C parental possui relações filogenéticas com a cepa brasileira. Esta conexão envolvendo o recombinante BC europeu a países sulamericanos, como o Brasil, também é mencionada por Simonetti e colaboradores, (2014). Os autores descrevem a CRF\_60, que também foi verificada na França (FRANGE et al., 2008), e os estudos evolutivos apresentados nestes trabalhos evidenciam amostras brasileiras em nós de interesse na topologia da árvore filogenética. Baseado nesta evidência, foram sugeridas duas hipóteses acerca do surgimento da CRF\_60: acredita-se que tenha surgido em território sulamericano e foi transmitida à Europa pelo transito de turistas; ou a CRF\_60 emergiu na Europa, induzida pelos vírus encontrados em viajantes destas localidades (SIMONETTI et al., 2014). Apesar do banco de dados do LANL apontar uma representante desta CRF no Brasil, não há artigo classificando esta amostra (JN642442) (Figura 9). Os autores do trabalho que gerou esta sequência identificaram dois recombinantes BC no estado de São Paulo, sendo uma delas a CRF\_31 (SANABANI et al., 2011).

A forma recombinante circulante descrita no Brasil é a CRF\_31 (SANTOS et al., 2006), identificada no estado do Rio Grande do Sul. Os primeiros casos envolvendo este recombinante é datado do início da década de 90 (SANTOS et al., 2007) embora as evidências filodinâmicas estimem que esta variante tenha surgido na década de 80: 1986 (FRITSCH et al., 2018) e 1988 (PASSAES et al., 2009). A capital do estado, Porto Alegre, parece ter sido o ponto de origem da CRF\_31 (FRITSCH et al., 2018). De acordo com os autores, algumas características, como a alta densidade populacional, co-circulação dos subtipos parentais, altas taxas de HIV positivo e a associação ao uso de drogas injetáveis favoreceram o surgimento e rápida expansão deste recombinante até mesmo para fora da região Sul (SANABANI et al., 2011; FRITSCH et al., 2018). Passaes e colaboradores (2009) identificaram diferentes padrões entre amostras de vírus BC e relacionaram a manutenção da posição específica (B) existente na CRF\_31 em outras sequências do estudo, reforçando que o mecanismo de evolução entre cepas do subtipo puro e o recombinante BC, impulsionando a diversidade de novos padrões (recombinantes de segunda geração). Há estudos de epidemiologia molecular evidenciando a dispersão dos recombinantes BC para todo o território brasileiro: Sudeste (SANABANI et al., 2011), Centro-Oeste (FERREIRA et al., 2011), Norte (COSTA et al., 2016) e Nordeste (DELATORRE et al., 2017).

Figura 10 - Classificação da sequência JN692442 no banco de dados LANL

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS INFO search site Search

Make Tree Download Sequences Save Background Info Make Histogram Geography Clear

Displaying 1 - 1 of 1 sequences found:

Select all Unselect all Invert selection Select record to List 1 records per page

Click on field name to sort in ascending or descending order

#	Select	Patient Code	Accession Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	<input type="checkbox"/>	Blast 2040(104149)	JN692442	02BR2040_60_BC	BRAZIL	2002		8934	HIV-1

last modified: Thu Oct 28 12:26 2021

Fonte: <https://www.hiv.lanl.gov/components/sequence/HIV/search/search.comp>.

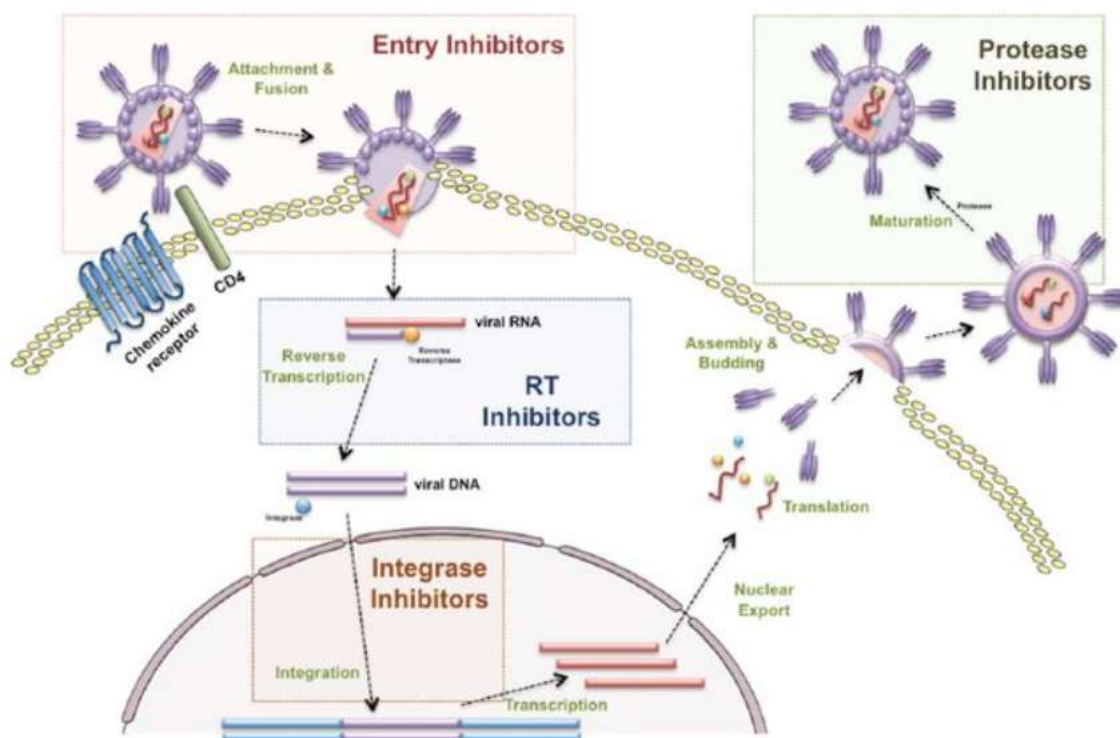
## 1.6 Repercussão da variabilidade genética do HIV-1 no tratamento e terapia antiretroviral

A ocorrência de mutações em regiões gênicas além de garantir a extensa diversidade existente pode ocasionar resistência a medicamentos. Atualmente a Terapia Antirretroviral (ART) é baseada em fármacos que atuam em alvos específicos da estrutura do vírus e acontece mais frequentemente em associação medicamentosa (SANTORO et al., 2013). Desde 1996 brasileiros vivendo com HIV/AIDS tem acesso gratuito ao tratamento que tem por objetivo proporcionar melhora na qualidade de vida, reduzindo o risco de mortalidade pela evolução da infecção até a síndrome da imunodeficiência, conservando as defesas imunológicas diminuindo a carga viral na circulação que, conseqüentemente, contribuem para o controle da infecção e epidemia (MACHADO et al., 2017; NOMAGUCHI et al., 2017; SANTORO et al., 2013).

Os alvos da terapia (Figura 10) são mais comuns ao nível de enzimas codificadas pelo gene *pol*. Estas drogas apresentam-se como inibidores da Transcriptase Reversa, que se subdivide em duas categorias: inibidores nucleosídicos (INTR) e não-nucleosídicos (INNTR). Há os inibidores da Integrase e Protease que interferem diretamente na replicação e no curso da própria infecção. Além dessas três possibilidades, são disponíveis também fármacos da categoria dos inibidores de fusão, estes agem bloqueando a interação entre as proteínas do envelope viral e os marcadores de superfície celular do linfócito T, por exemplo e por fim a classe de antagonistas do receptor de quimiocina CCR5. (TANG et al., 2012).



Figura 11 - Ciclo de replicação do HIV e os alvos dos antiretrovirais.



Fonte: CHERESHNEV et al., 2013.

Diferentes subtipos estão propensos a desenvolver determinadas mutações de resistência e isto se deve a características particulares de cada forma genética (SANTORO et al., 2013). As diferenças nas sequências de nucleotídeos configuram motivos estruturais variáveis devido ao número de transições e transversões, que confere essa variabilidade entre as cepas do HIV-1. Para o desenvolvimento de resistência às diversas classes de antiretrovirais poucas alterações são necessárias, por exemplo, mutação na posição 106 na estrutura da Transcriptase Reversa dos subtipos B e C conferem-lhes resistência à classe de medicamento INNR (BRENNER et al., 2003; GONZALEZ et al., 2004).

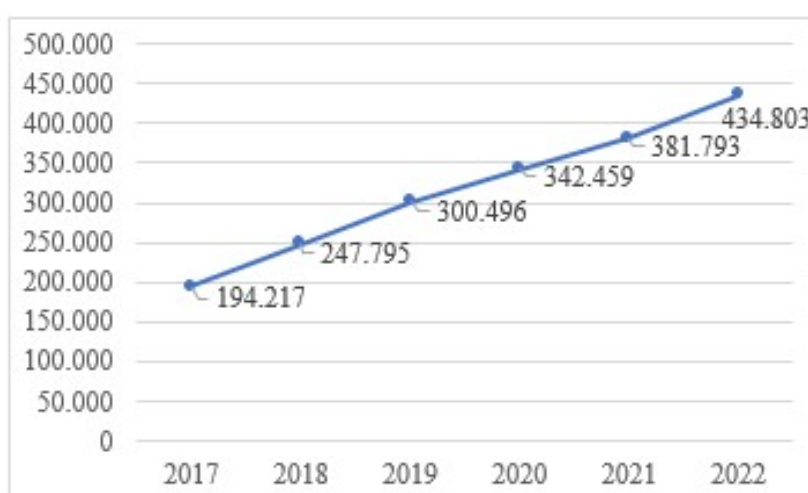
Segundo o estudo de Santoro e colaboradores (2013), diferentes mutações podem, também, surgir sob pressão da mesma classe de fármacos. A substituição de aspartato na posição 30 é comum em pacientes portadores do subtipo B sob uso de Nelfinavir, droga pertencente à classe de inibidores de Protease. A mesma droga quando utilizada em indivíduos dos subtipos C, F, G e CRF01\_AE apresentam prováveis substituições em outras posições como, na posição 90 troca de leucina para metionina e, também, asparagina para serina na posição 88 (GROSSMAN et al., 2014; SUSAKEM et al., 2008). A administração de Tipranavir induz a mutação de resistência (treonina para prolina na posição 74) mais frequentemente em

peças vivendo com HIV/AIDS do subtipo C em comparação aos pacientes que apresentam o subtipo B (SANTORO et al 2012). Os inibidores de entrada como, Enfuvirtide e Miraviroc têm como alvo a proteína gp41 e o receptor celular CCR5 respectivamente. (MELBY et al., 2006; SECLLEN et al., 2010). Alguns estudos revelaram que essas drogas parecem ser eficazes contra a maioria dos subtipos, porém, por possuir alvo no gene *env* (região altamente variável), pode haver polimorfismos que conferem resistência natural a esta classe de antiretroviral (AGHOKENG et al., 2005; POVEDA et al., 2004). A resistência aos antirretrovirais acontece em resposta ao acúmulo de mutações e o alto índice desses eventos nos diferentes subtipos de HIV-1 e este fato contribui para a dificuldade no desenvolvimento de vacinas e até novos recursos terapêuticos (SANTORO et al., 2013).

### 1.7 Justificativa

Segundo dados do Ministério da Saúde (2022) entre um período que corresponde a 15 anos (2007-2022), foram notificados 434.803 casos por HIV no Brasil. Apenas nos últimos seis anos (2017-2022), foram 240.586 novas infecções no país (Gráfico 1). De maneira geral, o perfil epidemiológico da infecção ainda é composta por indivíduos do sexo masculino, compreendendo 69,8%, contrastando com a taxa entre pessoas do sexo feminino, 30,2% e nesta população (sexo masculino) cerca de 52,9% têm idades entre 20-34 anos.

Gráfico 1 - Número de casos de HIV no Brasil por ano (2017 - 2021).

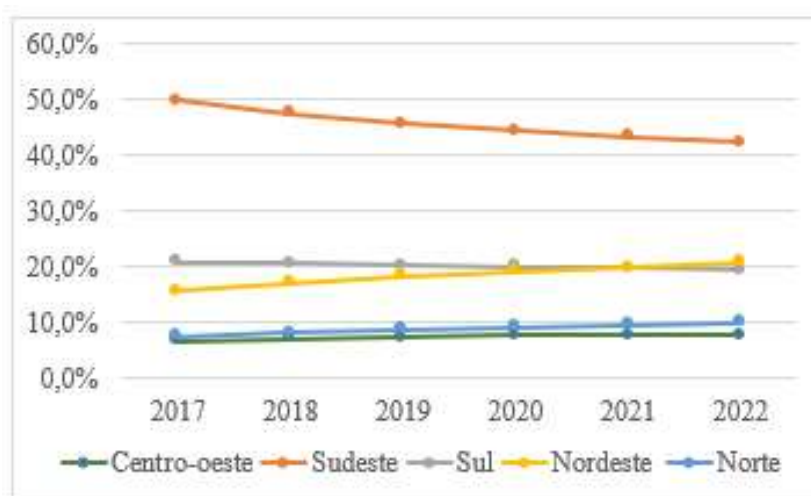


Fonte: adaptado boletim epidemiológico 2017 - 2022 (Ministério da Saúde).

Analisando estes dados através da divisão geográfica do Brasil, até o ano de 2020 o Sudeste apresentou 44,4% do número total de notificações por este vírus, em segundo lugar a

região Sul (20,0%), em seguida o Nordeste do país (19,8%), e por fim, as regiões Norte (9,5%) e Centro-oeste (7,7%) (MINISTÉRIO DA SAÚDE, 2020). Esta distribuição têm se mantido ao longo dos últimos anos, com excessão dos anos de 2021 e 2022, a região Nordeste saltou da terceira para a segunda posição em número de casos (Gráfico 2). Embora seja possível verificar um discreto aumento nos números do Ministério da Saúde relacionados ao Nordeste, ao longo de cinco anos, não podemos descartar a possibilidade de subnotificações nesta plataforma, uma vez que o ano de 2020 foi atípico, devido ao efeito da pandemia de SARS-CoV-2.

Gráfico 2 - Número (%) de casos de HIV por região brasileira



Fonte: adaptado boletim epidemiológico 2017 - 2022 (Ministério da Saúde).

As regiões Sudeste e Sul são onde estão concentrados a maior parte dos estudos relacionados a genética e diversidade do HIV, por terem sido o centro da epidemia inicial da AIDS durante a década de 80 (MACHADO et al., 2017). A região Nordeste é composta por 9 estados (Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe) e aparece em segundo lugar com maior número de ocorrências por HIV, somando 75.618 eventos, que equivale a 20,7% (MINISTÉRIO DA SAÚDE, 2022).

A epidemiologia molecular de HIV-1 no mundo revela que quase metade das infecções são causadas pelo subtipo C (~47%) em seguida, pelo subtipo B (~12%) e o subtipo A (~10%) (HEMELAAR et al., 2018). O genótipo C é prevalente em países africanos, especialmente do sul e leste do continente, China e Índia. (HEMELAAR et al., 2018). A co-circulação de variantes prevalentes favorecem o surgimento de formas recombinantes entre B e C, Atualmente no banco de dados do Laboratório Nacional de Los Alamos são listadas 118 padrões de CRFs, 14 são cepas BCs (11%)

(<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>), três delas foram descritas apenas no ano de 2021 (CRF108\_BC, CRF110\_BC e CRF118\_BC) (Cañada et al., 2021; Gao et al., 2021; Li et al., 2021).

De fato, a epidemiologia molecular do HIV-1 no Brasil já é conhecida, o subtipo B é prevalente e após este, estão os subtipos C, F (F1), as formas recombinantes BC e BF1 (ARRUDA et al., 2018). Em menor número, são encontrados os subtipos D e recombinantes AG e o complexo, BCF1 (GRAF et al., 2021). O subtipo B é dominante, no território brasileiro, 25 de 27 estados tem maior prevalência por este subtipo, e dois estados da região Norte revelam taxas superiores a 90% de infecções por este subtipo, enquanto os maiores índices de infecção pelo subtipo C é restrito à região Sul, com apenas 2 (Rio Grande do Sul e Santa Catarina) de 27 estados apresentando altas de infecção por este subtipo cerca de 40 e 60% respectivamente. O subsubtipo F1 é amplamente distribuído, sendo detectado em 13 de 27 estados brasileiros, mantendo taxas de prevalência por volta de 10% e as regiões Sudeste, Norte e Nordeste revelam maiores índices de infecção por este genótipo (GRAF et al., 2021).

A dispersão do subtipo C, forma genética prevalente na região Sul, para os estados ao norte do Brasil já foi constatada (DELATORRE et al., 2013; GRAF et al., 2015), porém até a região Nordeste não havia sido evidenciada. Os trabalhos publicados nos estados da região Nordeste que realizaram o sequenciamento de genes do vírus, revelam um discreto aumento da prevalência deste subtipo ao longo dos anos: 1,7% (2011) até 5% (2017), com a descrição de recombinantes BC apresentando taxas de prevalência equivalentes ao subtipo C parental (até 4%). Estudos de diversidade genética que avaliaram o genótipo F1 também evidenciaram um aumento das infecções por esse subsubtipo 2,3% em 2002 até 10,6% em 2018 (AMARAL et al., 2017; ARRUDA et al., 2011; ARRUDA et al., 2018; DELATORRE et al., 2017; MONTEIRO-CUNHA et al., 2009; MONTEIRO-CUNHA et al., 2017; MOURA et al., 2015).

Estudos moleculares recentes evidenciaram um aumento no número de casos de HIV-1 positivos principalmente nas regiões Nordeste e Sudeste, as estimativas revelam que desde 2014 esta discreta elevação na prevalência pelas cepas não-B esteja ocorrendo, promovidas por fatores sociodemográficos, que estabelecem uma nova dinâmica de transmissão viral na população (GRAF et al., 2021; SOUTO et al., 2021). Análises de diversidade molecular confirmam que o subtipo C, subsubtipo F1 e o recombinante BC, de fato, têm sua frequência aumentada no país, incluindo na região Nordeste, os autores destacam que esse avanço pode estar relacionado às suas propriedades biológicas (GRAF et al., 2021; SOUTO et al., 2021). As pesquisas de epidemiologia molecular que dizem respeito ao estado da Bahia foram realizadas

com fragmentos do gene *pol*, a nível da Protease e Transcriptase reversa (DELATORRE et al., 2017; MONTEIRO-CUNHA et al., 2009; MONTEIRO-CUNHA et al., 2017). Uma vez que é frequente a ocorrência de recombinação entre subtipos distintos nos isolados virais do HIV, sobretudo em regiões onde mais de um subtipo é predominante (como ocorre no Brasil), a obtenção de novos genomas totais e parciais dos subtipos do HIV-1 permite um delineamento mais fidedigno da história evolutiva destes subtipos.

Além disso, novas análises contendo mais amostras dos subtipos C, F1 e BC permitem a investigação de características moleculares previamente associadas com a dispersão destes subtipos nos isolados circulantes na Bahia, em comparação com as demais formas genéticas. Sendo assim, este estudo contribui com a vigilância epidemiológica deste vírus, permitindo conhecer a sua variabilidade genética a partir das ferramentas de bioinformática baseando-se em técnicas dos estudos de filogenia, filodinâmica, filogeografia, buscando de assinaturas em aminoácidos, possibilitando a compreensão das particularidades moleculares dos isolados em circulação e o comportamento viral em razão do genótipo.

### 1.7.1 Hipótese

Os isolados do subsubtipo F1 encontrados no Brasil são filogeneticamente relacionados às cepas da África ocidental (GUIMARÃES et al., 2009). As cepas virais deste subsubtipo circulantes no estado da Bahia são relacionadas às cepas encontradas em outras regiões do país... Os vírus pertencentes ao subtipo C identificadas no estado da Bahia e na região Nordeste, são filogeneticamente relacionadas às que circulam na região Sul, onde foi identificado pela primeira vez (BELLO et al., 2012; DELATORRE et al., 2013). As amostras deste subtipo identificadas no estado da Bahia apresentam características moleculares específicas relacionadas com uma menor dispersão viral. Os estudos de diversidade genética apontam os genótipos mais prevalentes no território brasileiro (B, C e F1) (GRAF et al., 2021), a circulação de mais um subtipo em uma população favorece o aparecimento de novas formas recombinantes, deste modo, os novos padrões de recombinação entre os subtipos B e C encontrados na região Nordeste são diferentes dos identificados nas outras regiões do país e em outras localidades no mundo, onde este recombinante também é prevalente. As análises de bioinformática, associadas à obtenção de novas amostras de nucleotídeos virais podem fornecer mais evidências para responder estas questões.

## 2 OBJETIVOS

### GERAL

Investigar a origem e variabilidade molecular de isolados dos subtipo C, subsubtipo F1 e recombinante BC do HIV-1 circulantes na região Nordeste do Brasil.

### ESPECÍFICOS

- Obter novas amostras de isolados virais do subtipo C, subsubtipo F1 e recombinante BC do HIV-1 circulante no estado da Bahia;
- Investigar a origem e história evolutiva do subsubtipo F1 e do subtipo C do Vírus da Imunodeficiência Humana tipo 1 circulantes na região Nordeste do Brasil;
- Verificar os padrões de recombinação de isolados BC do HIV-1 encontrados no estado da Bahia e comparar com amostras do Brasil e do mundo disponíveis em banco de dados públicos;
- Investigar características moleculares nos vírus do subtipo C circulantes no estado da Bahia e Brasil que possam estar relacionadas com a dispersão e patogenicidade comparando com sequências do mundo.
  - Predizer a utilização dos receptores de quimiocinas pelos isolados do subtipo C, com base na sequência de aminoácidos.
  - Buscar pelo dipeptídeo di-cisteína ou cisteína-serina nas posições nas 30 e 31 na proteína traduzida a partir dos nucleotídeos do gene *tat*.
  - Avaliar a presença dos motivos GSXXS e WL na proteína *nef*.
  - Verificar a presença do motivo 483L-484Y, mutações ou polimorfismos na região GAG-protease

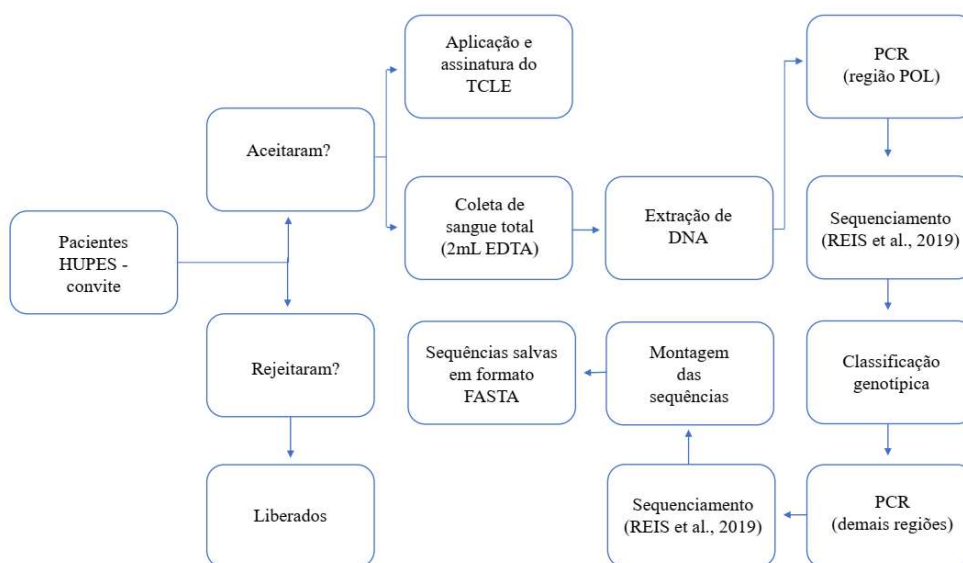
### 3 MATERIAL E MÉTODOS

#### 3.1 Obtenção de genomas completos de isolados virais do subtipo C do HIV-1 circulante no estado da Bahia

##### 3.1.1 População e Desenho de Estudo

O referido projeto é um estudo descritivo de corte transversal com amostragem de conveniência que avaliou amostras sanguíneas de pessoas vivendo com HIV/AIDS. A população de estudo foi formada por pessoas de qualquer idade, apresentando sorologia positiva para HIV-1, que após assinarem termo de consentimento livre e esclarecido, responderam ao questionário epidemiológico e permitiram a coleta de 2 ml de sangue para o estudo. A amostra para o estudo foi obtida durante coleta sanguínea, no Laboratório de Pesquisa em Infectologia (LAPI) do Hospital Universitário Prof. Edgard Santos (HUPES) para realização de carga viral do HIV ou estudo das subpopulações linfocitárias. O questionário epidemiológico foi preenchido pelo pesquisador e equipe; esse documento contém informações sobre identificação, data de nascimento, gênero, data do diagnóstico do HIV-1, uso de Terapia Antirretroviral (TARV), provável forma (via) de infecção, orientação sexual e discordância sorológica dos parceiros atuais. O projeto e o termo de consentimento livre e esclarecido foram submetidos, ao Comitê de Ética em Pesquisa da Instituto de Ciências da Saúde e do HUPES (Parecer número 1.395.808). A amostra de sangue foi utilizada para a extração do DNA e esta foi utilizada para a obtenção das sequências genômicas virais. As sequências geradas foram, então, analisadas no Núcleo de Bioinformática do Instituto de Ciências da Saúde (ICS) na Universidade Federal da Bahia (UFBA) através de ferramentas de Bioinformática para a caracterização do subtipo viral, baseada inicialmente no gene *pol*. Foram obtidas 102 sequências nucleotídicas a partir coleta de amostras dos pacientes do Hospital Universitário. Quatro amostras possuíam genotipagem prévia (1 do subtipo C, 1 do subsubtipo F1 e 2 recombinantes BF1) e 98 não tinha informação de genótipo. Das 98 sequências, 73 foram identificadas como subtipo B (74,5%), 16 recombinantes BF1 (16,3%), 4 subtipo C (4,1%), 3 subsubtipo F1 (3,1%), 1 subtipo D (1,0%), 1 recombinante BC (1,0%), 1 recombinante A1B (1,0%) e 1 recombinante CRF02\_AG/BF1 (1,0%). Os genomas obtidos das cepas C, F1, BC e BF1 foram editados e submetidos as análises de recombinação, investigação das relações filogenéticas, dos perfis mutacionais de resistência aos antirretrovirais e busca de assinaturas moleculares e funcionais (Figura 11).

Figura 12 – Fluxograma de recrutamento, coleta de amostras das pessoas vivendo HIV/AIDS acompanhados pelo Hospital Universitário Professor Edgard Santos e sequenciamento.



Fonte: autoria própria.

### 3.1.2 Coleta do sangue

As amostras de sangue foram obtidas por venopunção asséptica em tubos monovetes contendo anticoagulante EDTA (Ácido Etileno Diamino Tetracético). Foram coletados dois mililitros (2 mL) de sangue total de cada participante, os quais foram armazenados em freezers específicos a  $-20^{\circ}$  C.

### 3.1.3 Obtenção de genoma proviral

A extração e purificação do DNA foi realizada a partir do sangue total dos pacientes com PureLink™ Genomic DNA Mini Kit (Invitrogen™) de acordo com orientações do fabricante. A quantificação do DNA extraído foi realizada utilizando o NanoDrop (Thermo Fisher Scientific). As sequências genômicas virais foram obtidas a partir da Reação em Cadeia da Polimerase (PCR) em Termociclador MasterCycler Gradient Thermal Cycler (Eppendorf) utilizando-se um protocolo publicados anteriormente (REIS et al., 2019). Este protocolo estabelece a estratégia de *nested* PCR para quatro regiões sobrepostas aqui denominadas GAG, POL, ENV e LTR, correspondendo aos genes estruturais do HIV-1 (REIS et al., 2019). Para o primeiro fragmento amplificado, denominado GAG (5'LTR-gag-pro), foram utilizados os primers SCAOSD (GGGACTTTCCGCTGGGGACTTTC) e LR51



(GTATTCCTAATTGAACYTCC) como iniciadores externos e SCANS (CGAGCCCTCAGATGCTGCATATAAGC) e DP11 (CCTGGCTTYAATTTTACTGGTA) como iniciadores internos, resultando na amplificação de um fragmento de 2186 pb correspondente aos nucleotídeos 408 a 2594 referentes à cepa de referência HXB2. Para o segundo fragmento amplificado, denominado POL, foram utilizados os primers DP10 (CAACHCCCTCTCMGAAGCAGGAGC) e SCCNAS (TGTCCTGTAATAAACCCGAAAATTTTG) como iniciadores externos e DP16 (CCTCARRTCACTCTTTGGCARC) e SCCOAS (TATTCTTTCCCCTGCACTGTA) como iniciadores internos, resultando na amplificação de um fragmento de 2567 pb correspondente aos nucleotídeos 2253 a 4830 referentes à cepa de referência HXB2. Já para o terceiro fragmento amplificado, denominado ENV (INT-gp120), foram utilizados os primers MMINT8 (GCAGGAAGATGGCCAGT) e ED14 (TGYCTGGAGCTGYTTRATGCCCCAGAC) como iniciadores externos e MMINT3 (CCCTACAATCCCCAAAGTCAAG) e ED12 (AGTGCTTCCTGCTGCTCCCAAG) como iniciadores internos, resultando na amplificação de um fragmento de 3160 pb correspondente aos nucleotídeos 4653 a 7811 referentes à cepa de referência HXB2. Para o quarto fragmento amplificado, denominado LTR (gp41-LTR), foram utilizados os primers ED5 (ATGGGATCAAAGCCTAAARCCATGTG) e SCDOAD (AGTCACACAACAGACGGGCACACAC) como iniciadores externos e JH44 (ACAGTRCARTGYACACATGG) e LTR2 (AGGCAAGCTTTATTGAGGCTTAAG) como iniciadores internos, resultando na amplificação de um fragmento de 2671 pb correspondente aos nucleotídeos 6954 a 9625 referentes à cepa de referência HXB2. As condições de ciclagem utilizadas para a PCR foram: 94°C/2 min; 35 ciclos (94°C/30s; temperatura de anelamento/30s; 72°C/2min); 72°C/2min; 4°C (REIS et al., 2019). Os amplicons gerados foram então purificados utilizando o kit PureLink® Genomic DNA Mini Kit (Invitrogen) de acordo com o protocolo do fabricante. Para verificar o sucesso da amplificação, os produtos foram submetidos à eletroforese em gel de agarose a 1% em tampão TAE (tampão Tris-Acetato-EDTA) e quantificados utilizando o Nanodrop (Thermo Scientific). Em algumas amostras foi necessário utilizar outros protocolos para obter sucesso na amplificação de regiões menores do genoma. Esses protocolos alternativos foram igualmente cedidos e parcialmente realizados no LABAIDS, coordenado pela Dr<sup>a</sup>. Monick Guimarães na Fiocruz-RJ.

#### 3.1.4 Sequenciamento

O sequenciamento das amostras, utilizando a metodologia de Sanger, foi realizado em parceria com a rede de Plataformas Tecnológicas da Fiocruz Bahia e Rio de Janeiro (<https://plataformas.fiocruz.br/>). Tal sequenciamento foi realizado utilizando a tecnologia de Sanger no termociclador PerkinElmer 9600, utilizando-se o kit BigDye Terminator Cycle Sequencing (Perkin Elmer Applied Biosystems Division, Foster City, CA). A termociclagem de 25 ciclos de 96°C (10 s), 50°C (5 s), e 60°C (4 min). Os produtos foram então precipitados com isoetanol (70%) e posteriormente desnaturados a 96°C por 2 minutos, sendo então sequenciados no sequenciador automático Perkin-Elmer/ABI 3100 (Applied Biosystems, Foster City, CA) de acordo com as instruções do fabricante.

### 3.1.5 Edição dos Genomas

As seqüências geradas foram analisadas no Núcleo de Bioinformática, localizado no Instituto de Ciências da Saúde – Universidade Federal da Bahia. As *contigs* foram concatenadas através da sobreposição de suas extremidades utilizando o programa Geneious (KEARSE et al., 2012), norteado pela seqüência consenso do HIV-1 (HXB2 número de acesso: K03455) disponível no banco de dados do Los Alamos (LANL, do inglês *Los Alamos National Laboratory*) e manualmente editadas de acordo baseado nos eletroferogramas para a obtenção de uma seqüência única e completa de cerca de 8.000 pares de base (pb).

### 3.1.6 Determinação do subtipo das amostras

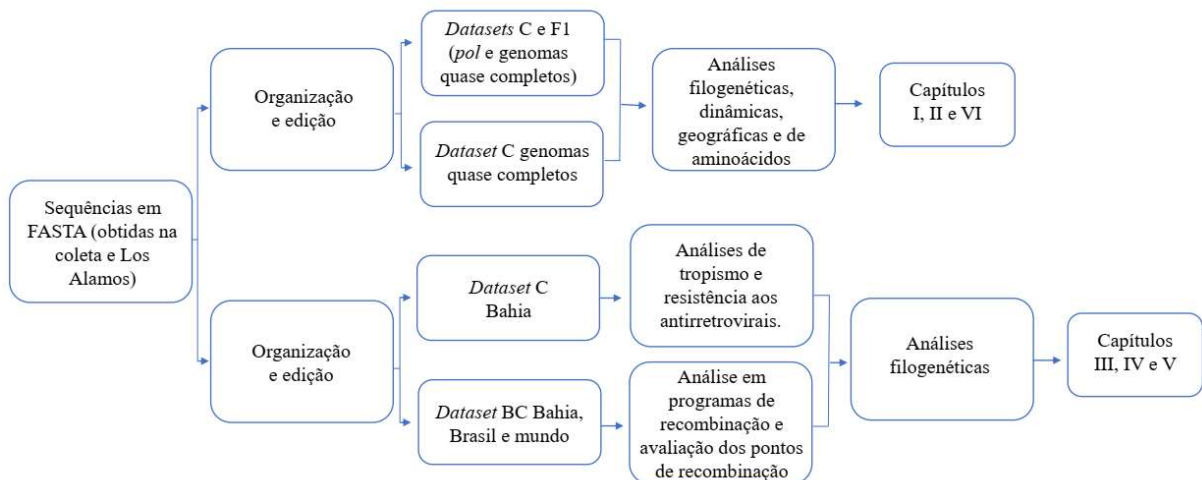
Todas as novas seqüências obtidas foram analisadas no programa BLAST (*Basic Local Alignment Search Tool*) para Reconhecimento de possíveis contaminações (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). As seqüências geradas foram submetidas às análises filogenéticas para a determinação dos subtipos do HIV-1. Estas seqüências foram alinhadas junto com 42 amostras de referência de todos os subtipos do grupo M do HIV-1 obtidas no banco de dados de Los Alamos (LANL (<https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>)) utilizando-se o programa MAFFT on-line (KATOY et al. 2005).

## 3.2 Busca de seqüências nucleotídicas e banco de dados

As seqüências genômicas foram obtidas a partir da base de dados do LANL

(<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) (Figura 12). Foram aplicados filtros de seleção, para o subtipo C, a busca foi restringida ao gene *pol*, selecionando o país de amostragem (Brasil). Para o subsubtipo F1, duas buscas foram realizadas, gene *pol* e sequências de genoma com comprimento quase completo, do inglês NFLG (*Near-Full length Genome*); para as amostras recombinantes BC, a pesquisa foi direcionada à seleção de amostras NFLG sem distinção de localidade. Todas as sequências estão catalogadas em planilhas do Microsoft Office Excel 2020, e esta etapa norteou a identificação, renomeação e inserção de informações adicionais sobre análise do material publicado nos artigos, de modo a definir os critérios de inclusão das sequências de nucleotídeos desta tese: apresentar informação de local e ano de obtenção das amostras. Os critérios de exclusão foram: sequências que apresentavam altas taxas de lacunas, bases degeneradas, códons de parada, registros duplicados, e sequências clones.

Figura 13 – Fluxograma das análises com as sequências de nucleotídeos obtidas no Hospital Universitário Professor Edgard Santos e no banco de dados do Laboratório Nacional Los Alamos.



Fonte: autoria própria.

### 3.3 Análises das sequências de nucleotídeos e análises filogenéticas

A montagem e edição dos *datasets* foram realizadas através do programa BioEdit (HALL 1999) e o alinhamento múltiplo na plataforma MAFFT on-line (KATO et al., 2005). O sinal filogenético dos conjuntos de dados foram investigados com auxílio da ferramenta *TreePuzzle* (SCHMIDT et al., 2003), para as análises filogenéticas utilizamos o *IQTree webservice* (TRIFINOPOULOS et al., 2016) e *IQTree local* (NGUYEN et al., 2015). Os

modelos evolutivos foram estimados no Modelfinder, implementado no IQTree (KALYAANAMOORTHY et al., 2017). Os suportes estatísticos foram estimados pelas técnicas do aLRT (*approximated Likelihood Ratio Test*) e *bootstrap* em 1000 reabordagens (ANISIMOVA & GASCUELA et al., 2006). As árvores filogenéticas geradas foram visualizadas utilizando o programa FigTree versão 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). Por fim, as seqüências geradas neste trabalho serão enviadas ao banco de dados GenBank (<http://www.ncbi.nih.gov/Genbank/>). A predição de utilização do coreceptor ocorreu através da plataforma on-line Geno2pheno (<https://www.geno2pheno.org/>).

### 3.4 Análise de recombinação

No intuito de obter um melhor entendimento da epidemia de AIDS no Nordeste, o comportamento filogenético das seqüências dos vírus recombinantes envolvendo o subtipo BC, foram comparado com o de seqüências de genoma quase total de vírus recombinantes provenientes de outras regiões brasileiras e de outras localidades do mundo. As seqüências foram avaliadas a partir do JphMM (*Jumping Profile Hidden Markov Model*), disponível no servidor on-line ([http://jphmm.gobics.de/submission\\_hiv.html](http://jphmm.gobics.de/submission_hiv.html)) e os mosaicos contendo os padrões de recombinação gerados a partir desta análise foram salvos em formato PDF. O jpHMM utiliza algoritmos baseados no modelo oculto de *Markov* que comparam a seqüência analisada com alinhamentos múltiplos de seqüências de diferentes subtipos de HIV-1. A partir do alinhamento de cada subtipo é criado um modelo probabilístico com o qual a seqüência será comparada. No entanto, diferentes fragmentos da seqüência analisada são alinhados localmente aos diferentes subtipos, permitindo assim a identificação de recombinantes com maior facilidade e rapidez (MELO, 2011). Para evitar viéses de metodologia, foram utilizados ainda os programas RDP5 (*Recombination Detected Program - 5*) (MARTIN et al., 2021), que utiliza uma associação de diferentes metodologias e o SimPlot, que submete as seqüências à análise de recombinação através do método *bootscanning* implementado no programa SIMPLOT (SALMINEN et al., 1995). Este método consiste na aplicação de técnica de *bootstrap* (reamostragem aleatória) em segmentos consecutivos do genoma [por exemplo, segmentos de 300 pb a cada 20 pb (valor do passo)]. Cada amostra-teste é analisada individualmente através da comparação do grau de relação (expresso como valor de *bootstrap*) com amostras de referência de todos os subtipos e principais CRFs pertencentes ao recombinante BC. O resultado

é um gráfico onde o eixo X significa posição na seqüência, e o eixo Y significa valores de *bootstrap* para cada segmento. Cada ponto significa um novo segmento que é movido pelo valor do passo. Em cada segmento, a amostra de referência mais relacionada com a amostra analisada tem seu valor de bootstrap” plotado no eixo Y.

### 3.5 Determinação da história evolutiva e origem dos isolados do subtipo C e subsubtipo F1

Com o objetivo de investigar a origem e a relação filogenética dos isolados do HIV-1 circulantes as seqüências nucleotídicas obtidas foram organizadas em *datasets* distintos (subtipo C, região POL; subsubtipo F1 POL e NFLG) e foram submetidos à verificação do sinal temporal pelo TempEst 1.5.3. (RAMBAUT et al., 2016), e depois avaliados em programas que utilizam a estatística Bayesiana implementados no pacote BEAST 1.10.4 (DRUMMOND et al., 2007). Foram considerados suportados estatisticamente valores igual ou superior a 0,80 de Probabilidade Posterior (PP).

### 3.6 Análise das seqüências de aminoácidos

As seqüências nucleotídicas obtidas através do banco de dados Los Alamos foram aplicadas à ferramenta CD-HIT (HUANG *et al.*, 2010), utilizada para filtrar as seqüências, baseado na similaridades entre as amostras, aplicando o *cut-off* de 0,9. As amostras foram traduzidas em seqüências de aminoácidos através do programa GENEDOC (NICHOLAS et al., 1997) e analisadas no programa BIOEDIT (HALL, 1999) quanto a presença de assinaturas de aminoácidos características. Para a identificação dos motivos funcionais nas seqüências de aminoácidos, foi utilizado o Pacote de Informações Biológicas do banco de dados de proteínas *Prosite Database* disponível no site (<https://prosite.expasy.org>).

As seqüências foram então investigadas de acordo com a informação de ano de obtenção daquele genoma e quanto a presença de assinaturas de aminoácidos características e exclusivas do grupo de isolados do subtipo C circulantes no Brasil, na região Nordeste do Brasil, comparando com seqüências de diferentes regiões geográficas do mundo. Para isso fez-se necessária uma pesquisa adicional de seqüências do subtipo C a partir da ferramenta HIV-BLAST ([https://www.hiv.lanl.gov/content/sequence/BASIC\\_BLAST/basic\\_blast.html](https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html)) para a identificação de seqüências deste genótipo de outras localidades do mundo para relacionar ancestralidade das cepas C do país.

Para avaliar se as substituições estão sendo selecionadas positivamente ou negativamente no curso da evolução dos vírus, aplicamos nossos conjuntos de dados ao servidor DATAMONKEY (<https://www.datamonkey.org/>). Para evitar a contagem de situações em que mutações sinônimas forem saturadas, médias de todas as razões de pares dN/dS foram calculadas excluindo pares que produzam valores de dS superiores a 1. Para esse fim, será utilizado o teste de aproximação bayesiana rápida e sem restrições (FUBAR) (MURRELL et al., 2013). Este teste permite inferir a seleção difusa sítio-específica, a seleção diversificadora episódica na região de interesse e identificar a seleção episódica em sítios individuais. A diferença estatística positiva ou negativa da seleção será baseada no valor de P menor que 0,05.

### 3.7 Considerações éticas

Número da Comissão de Apresentação de Apreciação Ética (CAAE): 50261715.8.3001.0049 – aprovado pelo Comitê de Ética em Pesquisa (CEP) do Instituto de Ciências da Saúde da Universidade Federal da Bahia (UFBA) e pelo CEP do Hospital Universitário Professor Edgard Santos (HUPES/UFBA) localizado na cidade de Salvador, Bahia.

## 4 RESULTADOS

Os resultados desse trabalho estão descritos na forma de artigos científicos por capítulos. Todas as etapas das análises realizadas bem como a escrita foram conduzidas pelo autor Rodrigo Cunha Oliveira, em parceria e/ou orientação dos demais coautores dos manuscritos.

**Capítulo I:** Rastreamento a relação entre as cepas F1 do subsubtipo de HIV-1: uma perspectiva filodinâmica (artigo: *Tracing the relationship among HIV-1 sub-subtype F1 strains: a phylodynamic perspective*).

Link: [Memórias do Instituto Oswaldo Cruz - 0109\\_Tracing the relationship among HIV-1 sub-subtype F1 strains: A Phylodynamic perspective \(fiocruz.br\)](https://www.fiocruz.br/memorias/0109-Tracing-the-relationship-among-HIV-1-sub-subtype-F1-strains-A-Phylodynamic-perspective)

**Capítulo II:** Dispersão dinâmica do subtipo C do HIV-1 em direção a região Nordeste Brasileira. (artigo: *Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region*).

Link: [Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region | AIDS Research and Human Retroviruses \(liebertpub.com\)](https://www.liebertpub.com/doi/10.1089/aid.2017.0000)

**Capítulo III:** Análise molecular e filogenética do HIV-1 subtipo C no estado da Bahia, Nordeste do Brasil. (artigo: *Molecular and phylogenetic analysis of HIV-1 subtype C in Bahia state, northeastern Brazil*).

Link: [Molecular and phylogenetic analysis of HIV-1 subtype C in Bahia state, northeastern Brazil. | AIDS Research and Human Retroviruses \(liebertpub.com\)](https://www.liebertpub.com/doi/10.1089/aid.2017.0000)

**Capítulo IV:** Uma nova Forma Recombinante Circulante (CRF) do HIV-1 B/C detectada em três regiões brasileiras. (artigo: *A new HIV-1 B/C Circulating Recombinant Form (CRF) detected in three Brazilian regions*).

**Capítulo V:** Caracterização genômica de vírus recombinantes HIV-1 BC (artigo: *Genomic characterization of HIV-1 BC recombinant Viruses*).

**Capítulo VI:** Predição de características moleculares no subtipo C do HIV-1 encontrados no Brasil e em várias partes do mundo.

# Capítulo I

---

## Rastreado a relação entre as cepas F1 do subsubtipo de HIV-1: uma perspectiva filodinâmica

(artigo: *Tracing the relationship among HIV-1 sub-subtype F1 strains: a phylodynamic perspective*).

O HIV-1F1 circula em três continentes: África, Europa e América do Sul. No Brasil, esse subsubtipo circula com os genótipos (B e C) e diversas formas recombinantes, principalmente variantes do BF1. Duas cepas principais do subsubtipo F1 estão circulando em todo o mundo. A primeira (F1.I) foi encontrada entre amostras brasileiras (75%) e a segunda (F1.II) entre romenos (62%) e outros isolados europeus e africanos. A epidemia do subtipo F1 no Brasil originou-se de uma única entrada no país por volta de 1970. Essa amostra ancestral está relacionada a amostras isoladas em países europeus (França, Finlândia e Bélgica), possivelmente de origem africana. Além disso, foi observada migração adicional de cepas do Brasil para a Europa. Curiosamente, todos os diferentes padrões de BF recombinantes encontrados, mesmo aqueles de fora do Brasil, apresentam a mesma linhagem F1 (F1.I) como ancestral, o que pode estar relacionado à aquisição de vantagens adaptativas para as progênies recombinantes. Estes achados são importantes para o entendimento da origem e dinâmica do subsubtipo F1 e consequente melhor e maior compreensão da epidemia de HIV-1 F1 que ainda se espalha do Brasil para outros países.

*O referido artigo foi publicado na revista Memórias do Instituto Oswaldo Cruz.*

*Published in Volume 117. November 17, 2022. [http://doi.org/10.1590/0074-](http://doi.org/10.1590/0074-02760220109)*

*02760220109*



## Tracing the relationship among HIV-1 sub-subtype F1 strains: a phylodynamic perspective

Gabriela Porto Santos Almeida Silva<sup>1</sup>, Rodrigo Cunha Oliveira<sup>1</sup>, Juliana Sacramento Mota de Souza<sup>1</sup>, Marta Giovanetti<sup>2</sup>, Monick Lindenmeyer Guimarães<sup>3</sup>, Carlos Brites<sup>4</sup>, Joana Paixão Monteiro-Cunha<sup>1/+</sup>

<sup>1</sup>Universidade Federal da Bahia, Departamento de Bioquímica e Biofísica, Salvador, BA, Brasil

<sup>2</sup>Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Laboratório de Genética Celular e Molecular, Belo Horizonte, MG, Brasil

<sup>3</sup>Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de AIDS e Imunologia Molecular, Rio de Janeiro, RJ, Brasil

<sup>4</sup>Universidade Federal da Bahia, Faculdade de Medicina, Salvador, BA, Brasil

**BACKGROUND** The human immunodeficiency virus type 1, F1 sub-subtype (HIV-1 F1) circulates in three continents: Africa, Europe, and South America. In Brazil, this sub-subtype co-circulates with subtypes B and C and several recombinant forms, mainly BF1 variants.

**OBJECTIVES** This study aimed to reconstruct the dynamic history of HIV-1 F1 in Brazil.

**METHODS** HIV-1 near full-length genome and pol gene nucleotide sequences available in public databases were assembled in two datasets (POL671 and NFLG53) to cover the largest number of F1 sub-subtype sequences. Phylodynamic and temporal analyses were performed.

**FINDINGS** Two main strains of the F1 sub-subtype are circulating worldwide. The first (F1.I) was found among Brazilian samples (75%) and the second (F1.II) among Romanian (62%) and other European and African isolates. The F1 subtype epidemic in Brazil originated from a single entry into the country around 1970. This ancestral sample is related to samples isolated in European countries (France, Finland, and Belgium), which are possibly of African origin. Moreover, further migration (1998 CI: 1994-2003) of strains from Brazil to Europe (Spain and the UK) was observed. Interestingly, all different recombinant BF patterns found, even those from outside Brazil, present the same F1 lineage (F1.I) as an ancestor, which could be related to the acquisition of adaptive advantages for the recombinant progenies.

**MAIN CONCLUSIONS** These findings are important for the understanding of the origin and dynamics of the F1 sub-subtype and a consequent better and greater understanding of the HIV-1 F1 and BF epidemic that still spreads from Brazil to other countries.

Key words: HIV - F1 sub-subtype - phylodynamic - recombinants - BF1

The human immunodeficiency virus type 1 (HIV-1) infection is still a major public health problem worldwide with an estimated 38 million people living with HIV by the end of 2020. In the same year, 690,000 people died of HIV-related causes and nearly 1.5 million people were newly infected.<sup>(1)</sup> In Brazil, 342,459 new HIV-1 cases were reported in the last 13 years (2007 to June 2020), of which 41,909 occurred in 2019.<sup>(2)</sup> The control of viral infection by the host's immune system and the effectiveness of drugs and vaccines are greatly hampered by the genetic diversity of the virus. The HIV is classified into two types (HIV-1 and HIV-2); HIV-1 is the most prevalent worldwide, further divided into four groups (M, N, O, and P).<sup>(3)</sup> The M group has spread significantly, representing around 90% of all infections and diversifying

into subtypes, identified as A (sub-subtypes A1, 2, 3, 4, 6 and 7), B, C, D, F (sub-subtypes F1 and F2), G, H, J, K, and L.<sup>(4,5)</sup> Additionally, different strains can recombine generating intersubtype forms that can be classified as Unique Recombinants Forms (URFs) or Circulating Recombinant Forms (CRFs) when identified by near full-length genome (NFLG) sequencing in at least three non-epidemiologically related individuals.<sup>(6)</sup> Currently, the Los Alamos National Laboratory (LANL) lists 121 HIV-1 CRFs (<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

The different genetic variants of HIV have a heterogeneous distribution around the world. A recent systematic review estimated a global prevalence of 46.6% for subtype C, 12.1% for subtype B, 10.3% for subtype A, and 0.9% for subtypes F, H, J, and K. In the same study, URFs and CRFs corresponded to approximately 22.8% of HIV infections globally from 2010 to 2015<sup>(7)</sup>, and 9.6% of these were caused by URFs described in Latin America.<sup>(4)</sup>

The HIV-1 F1 was initially described by Dumitrescu et al.<sup>(8)</sup> in 1994, isolated from Romanian individuals. In Brazil, the first identification of the F1 strain occurred in four samples collected between 1989-1990, the same period that the virus was identified in Romania.<sup>(9,10)</sup> Later, the analysis of African sequences of the F subtype

doi: 10.1590/0074-02760220109

Financial support: CNPq, FAPESB.

MLG is a recipient of a fellowship from CNPq -305919/2018-0; RCO and JSMS are PhD fellowships (FAPESB).

+ Corresponding author: joana.monteiro@ufba.br

● <https://orcid.org/0000-0001-9074-457X>

Received 10 May 2022

Accepted 17 November 2022



allowed its classification into two sub-subtypes: F1 and F2.<sup>(11)</sup> Despite its low global prevalence, sub-subtype F1 is widely spread, being exceptionally prevalent in some specific countries of Central Africa, South America, and Europe.<sup>(12)</sup> Molecular epidemiology studies of HIV-1 F1 in Europe have been conducted, especially in Romania, where this subtype is epidemic and high prevalence rates were found among children and adults.<sup>(13)</sup> In South America, the first F1 sequences were described in Brazil<sup>(9,10)</sup> and Argentina<sup>(14)</sup>, through partial analysis of the envelope region. In the subsequent years, many partial genomic sequences of the F1 sub-subtype were collected from infected individuals in those two countries, but, after the analysis of larger genomic regions, most of the sequences were confirmed as BF1 recombinant viruses instead of pure F1 strains.<sup>(15)</sup> In fact, viruses with unique BF recombination patterns (URF) circulate in Brazil and more than half of the 19 BF circulating recombinant forms (CRF\_BF) were first identified in this country.

Phylogenetic studies allow the improvement of the understanding of viral history. In the present study, we used nucleotide sequences from worldwide to reconstruct the evolutionary movement of HIV-1 F1 in Brazil and other countries. This work gathers information on the biological diversity and genetic classification of sub-subtype HIV-1 F1 in Brazil. The generated data provide significant information for understanding the origin and dynamics of HIV-1 F1, which can be useful for the development of measures to control its dissemination.

#### MATERIALS AND METHODS

**Dataset** - To understand the transmission dynamics of the HIV-1 F1 sub-subtype in Brazil, nucleotide sequences of *pol* region and NFLG were retrieved from the HIV LANL (<http://www.hiv.lanl.gov>) and NCBI (<http://www.ncbi.nlm.nih.gov>) databases. All sequences available in the databases classified as F1 in the NFLG and *pol-protease/reverse transcriptase (pol-PR/RT)* region were collected. The *pol* fragment was chosen because it has the largest number of available sequences, due to its use in monitoring HIV drug resistance. Exclusion criteria of samples included the presence of recombination in the analysed region, presence of gaps, degenerate bases, undefined fragments, stop codons, missing information about sample collection year, sample location, and/or sequences with any major HIV-1 drug resistance mutations according to the Stanford database (<https://hivdb.stanford.edu/>, updated Oct 23, 2020). All collected sequences were renamed with a code of sample information (access number + country + city + year). From that, two datasets were built. The first dataset included 671 genomic sequences (POL671) covering a *pol* region corresponding to positions 2367-3161 (relative to HXB2 reference strain). Out of these, 662 were obtained from public databases and nine were obtained in the state of Bahia, northeastern Brazil through another ongoing study of our group: 4 F1 sequences (006, 015, 192, 203) and 5 BF1 sequences (059, 069, 109, 189, 200), which were classified as F1 throughout the analysed genomic fragment. Among the 662 previously identified sequences, 632 were pure F1 in the entire genome and 30

had fragments of subtype B or BF1 over other genomic regions. The POL671 dataset included sequences from Brazil and other 36 countries [Table and Supplementary data (Table I)], and the second dataset (NFLG53) comprised 53 NFLG sequences (positions 790-8898 relative to HXB2) from Brazil and other 12 countries [Table and Supplementary data (Table II)]. The NFLG sequences were chosen as they have the largest genomic representation, even with relatively fewer available sequences.

TABLE  
Origin of HIV-1 sub-subtype F1 isolates

Country	POL671(N)	NFLG53 (N)
Brazil	281	24
Romania	184	2
Spain	61	13
Italy	20	
Angola	19	3
Portugal	12	
Germany	9	1
Cape Green	8	
CD	8	
United States	7	
Austria	6	
Japan	6	
United Kingdom	6	3
Argentina	5	1
Belgium	5	1
Cyprus	5	1
China	4	
France	3	2
Turkey	3	
French Guyana	2	
Russian Federation	2	1
Bolivia	1	
Bulgaria	1	
Cameroon	1	
Canada	1	
CF	1	
Cuba	1	
Finland	1	1
Luxembourg	1	
Netherlands	1	
Nigeria	1	
Panama	1	
Poland	1	
Slovenia	1	
Sweden	1	
Zimbabwe	1	

N: number of sequences; CD: Democratic Republic of Congo; CF: Central African Republic.

*Sequence alignment and phylogenetic analysis* - Alignment was performed using the MAFFT online program (<http://mafft.cbrc.jp>)<sup>(16)</sup> under the command: `mafft-thread 8 - threadb 5 -threadit 0 -reorder -auto input> output` and manually edited using the BioEdit software.<sup>(17)</sup> Datasets were assessed for the presence of phylogenetic signals by applying the likelihood mapping analysis implemented in the TreePuzzle Program.<sup>(18)</sup> The percentage of fully resolved probability mapping quartets totals was 93.7% for pol dataset and 95.1% for NFLG dataset proving the phylogenetic signal for both tree reconstructions. The maximum likelihood (ML) trees were reconstructed using IQ-TREE 1.6.8 webserver.<sup>(19)</sup> Phylogenetic analyses were performed under the GTR+I+G nucleotide substitution model calculated in Modelfinder.<sup>(20)</sup> The reliability of each cluster was evaluated considering a bootstrap (BS) of 1,000 replicates and with an approximate likelihood-ratio test (aLRT) based on the Shimodaira-Hasegawa-like procedure.<sup>(21)</sup> The ML trees were visualised with FigTree program version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). The subtype was confirmed through phylogeny using the ML method against 11 HIV-1 reference sequences (A1, B, C, D, F2, G, and K) obtained from Los Alamos (<https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) and through the Hidden Markov Model method (jpHMM) to exclude possible events of recombination in the samples within the analysed fragment.<sup>(22)</sup>

*Phylogenetic and temporal analysis of HIV-1 subtype F1* - For phylogenetic studies, ML trees were used to regress root-to-tip genetic distances against sample collection dates using TempEst v 1.5.1 (POL  $r^2 = 0.5071$ ; NFLG  $r^2 = 0.6337$ ).<sup>(23)</sup> In both datasets (NFLG53 and POL671), the BEAST package software 1.10.4 was applied to verify the marginal likelihood values by estimating Path sampling and Sample Stepping-stone to define the model and molecular clock.<sup>(24,25)</sup> Through the standard nucleotide substitution model (HKY) and the uncorrelated relaxed clock and exponential model, the maximum clade credibility trees (MCC) were obtained. The ML phylogeny was used as a starting tree for Bayesian time-scaled phylogenetic reconstruction. In the analysis of the *pol* fragment, four runs of Markov Chain Monte Carlo (MCMC) with 300 million states each were computed, sampling every 30 million steps, that value of effective sample size (ESS) of proper mixing was 334.2 (> 200). The log and trees files were combined in Log-Combiner discarding 10% in each run (POL671 = 40%) and (NFLG53 = 20%) as burn-in; the convergence of the MCMC chains was verified using Tracer v.1.7.1.<sup>(26)</sup> The Bayesian coalescent inference of the population analysis was performed with the NFLG dataset based on the MCMC model and chain number already described.

## RESULTS

*Dataset characterisation* - The POL671 dataset comprised 39 sequences from Africa, 289 from South America, 11 from Central/North America, 17 from Asia, and 316 from Europe. The oldest sequence was obtained in 1989 and the most recent was collected in 2019. Re-

garding the NFLG53 dataset, among the 53 sequences, one was collected in Asia, three in Africa, 35 in South America, and 24 in Europe. The oldest sequence was collected in 1989 and the most recent in 2016.

*Phylogenetic inference* - The ML analysis of POL671 [Supplementary data (Fig. 1)] dataset shows a topology with two large nodes. The F1.I lineage supported by aLRT = 85/ BS = 92 includes sequences from 20 countries (Brazil, Spain, Germany, Portugal, Luxembourg, Italy, Argentina, Bolivia, United States, Japan, French Guyana, China, France, Finland, Belgium, United Kingdom, Germany, Cameroon, Cyprus, and Panama). Most of the samples in this upper cluster belong to South America, including all sequences from Brazil (n = 281). All 35 F1 sequences retrieved from BF1 recombinants sequences, regardless of origin, clustered inside the clade F1.I. Furthermore, only one African F1 sample (from the Republic of Cameroon aLRT 85/BS 99) was inside this group. The F1.II lineage grouped samples from 25 countries (Romania, China, Japan, Belgium, Spain, Canada, Angola, Portugal, Germany, Cape Verde, Nigeria, Democratic Republic of Congo, Russian Federation, France, Central African Republic, Cuba, Slovenia, Sweden, Italy, United Kingdom, Netherlands, Cyprus, Turkey, Poland, and Austria) with aLRT = 85/BS = 82. Notably, most sequences from African countries (97.4%) were in this cluster. The external group of this cluster is represented by five sequences from African countries (Angola, Democratic Republic of Congo, and Zimbabwe [aLRT 84/BS 99]).

Of note, in the NFLG analysis, the two main nodes were also present in the ML tree [Supplementary data (Fig. 2)] (F1.II node: aLRT 99/ BS 100 and F1.I: aLRT 99 / BS 100), confirming the separation between the strains circulating in Brazil from those found in Romania and Africa. The ancestor of the F1.I node seems to originate from two clusters: one with Brazilian sequences and a second cluster with European sequences only (from Finland, France, and Belgium).

To understand the transmission dynamics of HIV-1 F1 in Brazil, Bayesian analyses were conducted, and the topologies obtained from both datasets (Figs 1-2) were remarkably similar to the ML trees [Supplementary data (Figs 1-2)], showing the two major F1 lineages circulating worldwide. Again, Brazilian, and Romanian sequences were always grouped in distinct clusters. In the Bayesian tree of pol (Fig. 1), the F1.I cluster was composed of 374 sequences, of which 280 (74.5%) came from Brazil. The common ancestor of this cluster would have circulated in 1946 (CI:1928-1959). The F1.II cluster grouped 297 sequences, mostly from Romania (84.6%). All new Brazilian sequences from the state of Bahia were intermixed with previously described Brazilian F1 sequences inside the F1.I clade, with high PP support. As in the ML tree, all 35 BF recombinant sequences are also grouped inside this clade.

The NFLG Bayesian tree (Fig. 2) shows, as in POL671 tree, an F1.II node with European and African sequences and an F1.I node with Brazilian sequences. The estimated time to the most recent common ancestor (TMRCA)

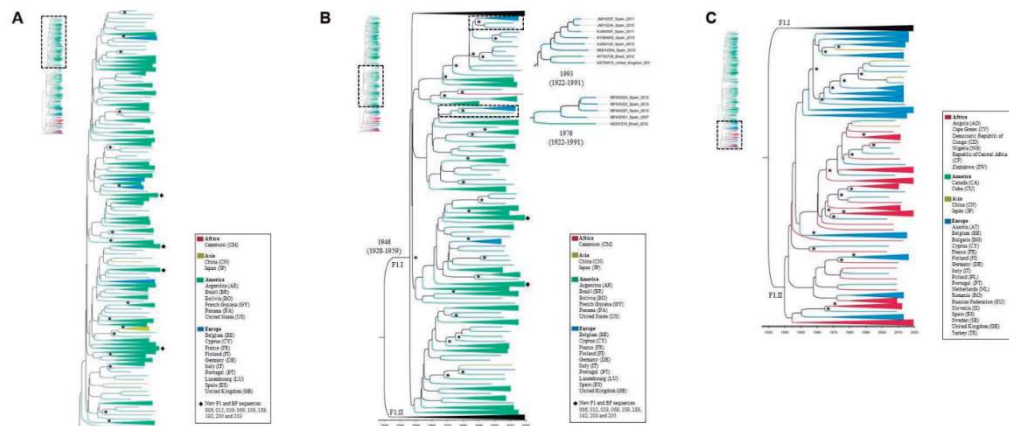


Fig. 1: time scale Bayesian maximum clade credibility trees (MCC) (n = 671) based on pol genomic fragment. Tree was reconstructed using 671 HIV-1 sub-subtype F1 viral isolates from worldwide. (A) with emphasis on the tree top (B) middle and (C) bottom. Sequences were isolated in Brazil (280), Europe (316) and Africa (39). Branches with PP support (> 0.80) are marked with a \*. The colours of the branches refer to the location indicated by the legend. New Brazilian sequences are marked with ♦. Relevant branches were expanded for better visualisation, exposing its dating and confidence interval. The cluster of Brazilian sequences is shown in Fig. 1A-B (F1.I lineage) and collapsed in 1C for better visualisation of the lower branch of the tree (F1.II lineage).

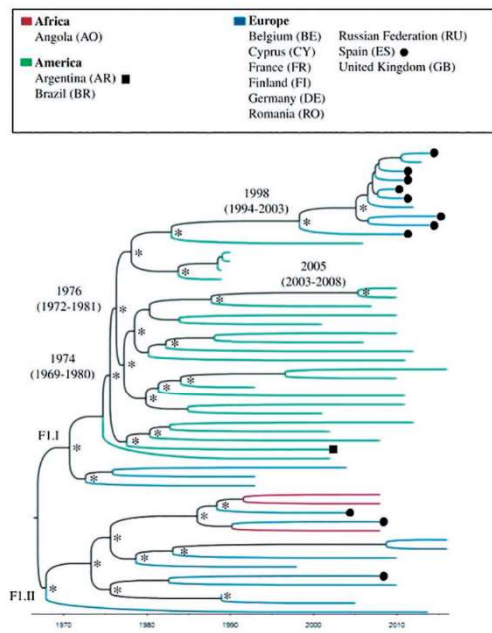


Fig. 2: time scale Bayesian maximum clade credibility trees (MCC) (n = 53) based on near full-length genome (NFLG) sequences showing the phylogenetic relationships among sub-subtype F1 viruses circulating in Brazil (25), Europe (24) and Africa (3). Branch colors represents the geographic region of origin, according to the legend. Important branches with PP support (> 0.80) are marked with \*. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating years. F1.I represents the cluster of Brazilian sequences.

of this Brazilian lineage is 1974 (CI: 1969-1980). This reconstruction also points to a subsequent migration of the F1.I lineage from Brazil to Europe (Spain and UK) in the late 1990s (1998 with CI 1994-2003 [PP = 1]).

*Bayesian coalescent inference of population analysis* - The Bayesian coalescent inference of population analysis from the NFLG dataset generated a population growth curve with a time interval between 1970 and 2015 (Fig. 3). This result shows the exponential growth of the F1 subtype until around 1995; from then on, there is a plateau between 1995 and 2005, followed by a decrease in the expansion of this subtype and a slight increase around 2009.

**DISCUSSION**

The two different datasets gave rise to ML and Bayesian trees with remarkably similar topologies. In all four trees [Figs 1-2, and Supplementary data (Figs 1-2)], one can observe the existence of a relationship between sequences from South American and European countries (F1.I node), more specifically between Brazil and Spain. Also, in the F1.II there is evidence of the phylogenetic link between samples from African and European countries; in this case, Romania is the most represented country. The F1 lineage in Brazil is more related to sequences from France, Finland, and Belgium, and these are close to F1 African sequences (mainly West and Central Africa).<sup>(27,28)</sup>

Furthermore, in the four analyses, all Brazilian sub-subtype F1 sequences formed a monophyletic group together with sequences from other regions (F1.I large group). This indicates a single entry of sub-subtype F1 in Brazil around the 70s, as shown in our Bayesian analyses (Figs 1-2). Earlier evolutionary analyses

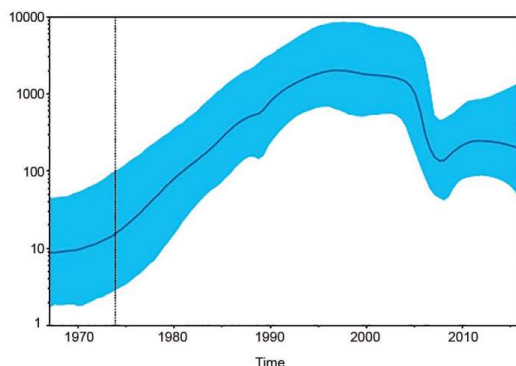


Fig. 3: Bayesian coalescent inference of population size of subtype F1 from dataset near full-length genome (NFLG)53 including 53 NFLG sequences (positions: 790-8898 relative to HXB2) from Brazil and others 11 countries (Angola, Argentine, Belgium, Cyprus, France, Finland, Germany, Romania, Russian Federation, Spain and United Kingdom). All sequences were obtained from the Los Alamos HIV sequence database ([www.hiv.lanl.gov/content/index](http://www.hiv.lanl.gov/content/index)).

revealed that the oldest isolates related to the F1 sequences were from the Democratic Republic of Congo and that their diversification occurred in that country in the late 1950s.<sup>(29)</sup> In Brazil, studies on the evolutionary history of the HIV F1 sub-subtype estimated that the onset of the epidemic occurred in the early 1980s.<sup>(30,31,32)</sup> In contrast to these findings, evaluating the tM-RCA data from our trees, it is notable that, in the analysis of the *pol* region, 12 out of 27 clusters involving Brazilian samples with  $PP \geq 0.80$  were dated around the 1970s, consistent with others.<sup>(12,15)</sup> In addition, the NFLG53 tree (Fig. 2) shows the probable introduction of HIV-1 F1 in Brazil by 1974 ( $PP = 0.98$ ). Our findings coincide with the end of Angola's independence war and the beginning of the civil war in 1975, which was followed by a wave of emigration and the participation of foreign troops from several countries.<sup>(27,33)</sup>

The HIV-1 F1 sub-subtype viral strain has been identified at the same time in Brazil and Romania. For this reason, a premeditated conclusion hypothesises an epidemiological and evolutionary relationship between the F1 viruses from these two countries. Despite of this, in all four phylogenetic trees of this study, the sequences from Romania and Brazil grouped into two distinct clusters. In the Brazilian monophyletic group, there are no sequences from Romania, as they were always grouped in the F1.II unlinked node. Banda et al.<sup>(34)</sup> in their phylogenetic analyses also showed that there is no epidemiological link between the two epidemics; Guimarães et al.<sup>(29)</sup> showed that the F1 strains from Romania and Brazil were grouped into two related but distinct clusters. In this context, this research reinforces previous findings with greater grounding and support for phylogenetic analyses.

The European subcluster inside the F1.I cluster in *pol* (Fig. 1B) and NFLG trees (Fig. 2) points to a recent relationship among Brazilian, Spanish, and English viruses. According to these reconstructions (ML and Bayesian),

the common ancestor of this European clade originated from the Brazilian cluster and dated from 1998 (CI 1994-2003), suggesting that there may be a contribution of the Brazilian F1 sub-subtype to the HIV-1 F1 epidemic in Europe, mainly Spain. This inference is supported by the recent spread of subtype F to new geographic areas (especially Spain) and its later expansion, showing a high rate of an effective number of infections in a short period.<sup>(35)</sup> Additionally, in the last decades, Spain has become one of the main destinations in Europe for immigrants and one of the western European countries with the highest rates of HIV infections.

The *pol* dataset holds 35 BF1 recombinants that were classified as F1 throughout the genomic fragment used in our analyses. All these recombinant sequences, including three non-Brazilian viruses (Japan, Bolivia, and Argentina), were inside the F1.I cluster, together with Brazilian F1 strains in both Bayesian (Fig. 1) and ML [Supplementary data (Fig. 1)] trees. It has been shown that BF1 recombinant sequences have many distinct recombination patterns, indicating different recombination events.<sup>(36)</sup> In the present study, despite the different recombination patterns observed among the 35 BF1 representatives, the clustering of all recombinants in the same monophyletic cluster suggests a common origin for these viruses. This may indicate that the pure F1 Brazilian ancestor could have arisen from a possible evolution after adaptive advantages associated with recombinant strains. Our results indicate that all BF1 recombinants in the present study have the same F1 lineage as the ancestral sequence, which could show a tendency of this ancestral lineage to recombine and the existence of adaptive advantages for the recombinant progenies.

The results of this study show a probable migration of the F1 sub-subtype from Africa to Europe and, after that, a single introduction of this strain to Brazil around the 70s (Figs 1A-B, 2). From these analyses, it is also possible to suggest a later migration from South America to Europe in 1998 (1994-2003) (Fig. 2). The consistency of these results is reflected by the greater number of sequences analysed, compared to earlier studies that focused on a single, smaller genomic region, and by the consistency of the topologies observed in *pol* and NFLG datasets in both Bayesian and ML analyses. These findings provide important information, involving molecular epidemiology, which contributes to understanding the origin, diversification, and dispersion dynamics of the F1 sub-subtype. This knowledge contributes to the surveillance and control of the HIV-1 F1 and BF1 epidemic that is still spreading from Brazil to South America and other European countries.

The Bayesian coalescent inference of population growth showed a temporal expansion of the F1 sub-subtype in the first decades of the epidemic, followed by the stabilisation of transmission. Lima et al.<sup>(32)</sup> compared a limited number of F1 samples from Pernambuco against F1 reference sequences between 1986 and 2010 and found that both datasets (Pernambuco and reference) were associated with a moderate growth phase followed by a significant increase in the frequency of HIV-1 F strains. Their plot also shows a plateau be-

tween 2005 and 2010, which corroborates our analysis, showing a slight decrease in the expansion of the F1 sub-subtype over the same period. Furthermore, Bello et al.<sup>(12)</sup> also showed the slow but constant growth of the F1 epidemic in adults, until the beginning of the 2000s; after that, the epidemic growth rate started to slow down. Based on temporal expansion analyses, our study included all HIV-1 F1 NFLG sequences available in the databases up to the year of collection. Despite of this, we suggest that further studies should be conducted to evaluate the expansion of the F1 sub-subtype over the next few years, based on the availability of new sequences. Furthermore, we believe that new temporal expansion analyses can be effective for a comparison between recombinant and pure F1 to achieve further interpretations of adaptive advantages.

#### AUTHORS' CONTRIBUTION

GPSAS and RCO performed the analysis and wrote the manuscript; JSMS sequenced new samples and wrote the manuscript; MG guided the phylodynamic analysis and revised the manuscript; MLG sequenced new samples and revised the manuscript; CRBA obtained new samples and revised the manuscript; JPMC conceived and planned the study, analysed and wrote the manuscript. All authors supplied critical feedback and revised the manuscript. The authors declare that they have no conflict of interest in this work.

#### REFERENCES

- UNAIDS.org.br. Brazil. Joint United Nations Program on HIV/AIDS. 2021. Available from: <https://unaids.org.br/estatisticas>.
- MS/SVS - Ministério da Saúde/Secretaria de Vigilância em Saúde. Boletim, Número especial. Boletim Epidemiológico HIV 2020. 2020. Available from: <http://www.aids.gov.br/pt-br/pub/2020/boletim-epidemiologico-hiv-aids-2020>.
- Korber B, Muldon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science*. 2000; 288(5472): 1789-96.
- Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Kirtley S, Gouws-Williams E, et al. Global and regional epidemiology of HIV-1 recombinants in 1990-2015: a systematic review and global survey. *Lancet HIV*. 2020; 7(11): e772-81.
- Yamaguchi J, Vallari A, McArthur C, Sthreshley L, Cloherty GA, Berg MG, et al. Brief report: complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *J Acquir Immune Defic Syndr*. 2020; 83(3): 319-22.
- Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV nomenclature proposal - a reference guide to HIV-1 classification. *Science*. 2000; 288(5463): 55-6.
- Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis*. 2019; 19(2): 143-55.
- Dumitrescu O, Kalish ML, Kliks SC, Banda CI, Levy JA. Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J Infect Dis*. 1994; 169(2): 281-8.
- Louwagie J, Delwart EL, Mullins JI, McCutchan FE, Eddy G, Burke DS. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res Hum Retroviruses*. 1994; 10(5): 561-7.
- Morgado MG, Sabino EC, Shpaer EG, Bongertz V, Brigido L, Guimaraes MD, et al. V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses*. 1994; 10(5): 569-76.
- Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N, et al. High diversity of HIV-1 subtype F strains in Central Africa. *Virology*. 1999; 259(1): 99-109.
- Bello G, Afonso JM, Morgado MG. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect Genet Evol*. 2012; 12(5): 1079-86.
- Apetrei C, Loussert-Ajaka I, Collin G, Letourneur F, Duca M, Saragosti S, et al. HIV type 1 subtype F sequences in Romanian children and adults. *AIDS Res Hum Retroviruses*. 1997; 13(4): 363-5.
- Marquina S, Leitner T, Rabinovich RD, Benetucci J, Libonatti O, Albert J. Coexistence of subtypes B, F, and as B/F env recombinant of HIV type 1 in Buenos Aires Argentina. *AIDS Res Hum Retroviruses*. 1996; 12(17): 1651-4.
- Aulicino PC, Bello G, Rocco C, Romero H, Mangano A, Morgado MG, et al. Description of the first full-length HIV Type 1 subtype F1 strain in Argentina: implications for the origin and dispersion of this subtype in South America. *AIDS Research Hum Retroviruses*. 2007; 23(10): 1176-82.
- Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005; 33(2): 511-8.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999; 41: 95-8.
- Schmidt HA, Petzold E, Vingron M, Haeseler AV. Molecular phylogenetics: parallelized parameter estimation and quartet puzzling. *J Parallel Distrib Comput*. 2003; 63(7-8): 719-27.
- 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(W1): w232-5.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017; 14(6): 587-9.
- Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol*. 2006; 55(4): 539-52.
- Schultz AK, Zang M, Bulla I, Leitner T, Kobber B, Morgenstern B, et al. jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res*. 2009; 37: w647-51.
- Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016; 2(1): vew007.
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7: 214.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29(8): 1969-73.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018; 67(5): 901-4.
- Laukkanen T, Carr JK, Janssens W, Liitsola K, Gotte D, McCutchan FE, et al. Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America. *Virology*. 2000; 269(1): 95-104.
- Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole E, Mulanga-Kabeya C, Nzilambi N, et al. Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. *AIDS Res Hum Retroviruses*. 2000; 16(2): 139-51.

29. Guimarães ML, Vicente ACP, Otsuki K, Silva RFFC, Francisco M, Morgado MG, et al. Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology*. 2009; 6: 39.
30. Bello G, Guimarães ML, Morgado MG. Evolutionary history of HIV-1 subtype B and F infections in Brazil. *AIDS*. 2006; 20(5): 763-8.
31. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SLM, et al. Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol*. 2007; 7(2): 263-70.
32. Lima K, Leal ES, Cavalcanti AMS, Salustiano DM, Medeiros LB, Silva SP, et al. Increase in human immunodeficiency virus 1 diversity and detection of various subtypes and recombinants in north-eastern Brazil. *J Med Microbiol*. 2017; 66(4): 526-35.
33. Op de Coul E, van der Schoot A, Goudsmit J, van den Burg R, Janssens W, Heyndrickx L, et al. Independent introduction of transmissible F/D recombinant HIV-1 from Africa into Belgium and The Netherlands. *Virology*. 2000; 270(2): 267-77.
34. Banea CI, Ramos A, Pieniazek D, Pascu R, Tanuri A, Schochetman G, et al. Epidemiologic and evolutionary relationships between Romanian and Brazilian HIV-subtype F strains. *Emerg Infect Dis*. 1995; 1(3): 91-3.
35. Gonzalez-Alba JM, Holguin A, Garcia R, Garcia-Bujalance S, Alonso R, Suárez A, et al. Molecular surveillance of HIV-1 in Madrid, Spain: a phylogeographic analysis. *J Virol*. 2011; 85(20): 10755-63.
36. Souza JSM, Silva Jr JJ, Brites C, Monteiro-Cunha JP. Molecular and geographic characterization of HIV-1 BF recombinant viruses. *Virus Res*. 2019; 270: 197650.

## Capítulo II

---

### Dispersão dinâmica do subtipo C do HIV-1 em direção à região Nordeste Brasileira.

(artigo: *Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region*).

O subtipo C é responsável por >50% das infecções pelo HIV-1 no mundo e atualmente é a forma viral predominante no Sul do Brasil. O subtipo C foi relatado em todas as regiões brasileiras; no entanto, a relação filogenética entre as cepas que circulam nessas regiões ainda não está clara. Este estudo teve como objetivo investigar a origem e a dispersão dinâmica do HIV-1 subtipo C em direção ao Nordeste do Brasil. Nossa análise filogenética sugere que a maioria das cepas do subtipo C circulantes no Brasil (99%) são descendentes da linhagem principal cuja entrada no país foi descrita anteriormente na década de 1970. De acordo com a literatura, foram relatadas introduções adicionais do subtipo C no país pela região Sudeste e neste estudo identificamos outro evento de entrada que ocorreu mais provavelmente pela região Norte. Além disso, nossa análise sugere que a disseminação do subtipo C para os estados do Nordeste brasileiro ocorreu por meio de múltiplas introduções independentes da linhagem principal que se originou no sul do Brasil entre meados dos anos 1980 e final dos anos 1990. Apesar da observação de eventuais novas introduções do subtipo C do HIV-1, nossos resultados destacam a predominância de uma única linhagem desse subtipo no Brasil e a importância da região Sul na sua disseminação pelo país.

*O referido artigo foi publicado na revista AIDS Research and Human Retroviruses.*

*Published in Volume: 37 Issue 12: December 15, 2021. <http://doi.org/10.1089/aid.2020.0308>*



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

Rodrigo Cunha Oliveira,<sup>1</sup> Tiago Gräf,<sup>2</sup> Filipe Ferreira de Almeida Rego,<sup>3</sup>  
 Gabriela Porto Santos Almeida Silva,<sup>1</sup> Marta Giovanetti,<sup>4,5</sup> and Joana Paixão Monteiro Cunha<sup>1,i</sup>

### 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

**I**NFECTION 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.<sup>1</sup> 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.<sup>2</sup> The intense genetic variability and capacity to establish reservoirs *in vivo* are the main obstacles in the fight against HIV/AIDS.<sup>3</sup> 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.<sup>4</sup>

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.<sup>5</sup> 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.<sup>6,7</sup> In the Southeast and Central West regions, B, C, F1, and BF1 cocirculate,<sup>8,9</sup> whereas in the South region, HIV-1 infections are caused mainly by subtypes B, C, and BC recombinants.<sup>9</sup>

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,<sup>10</sup> whereas most recent reports described a prevalence of 44%–66%,<sup>11</sup> pointing to an increase of subtype C prevalence in South Brazil over the last two decades, as previously reviewed.<sup>12</sup> Furthermore, the subtype C prevalence has also been reported in Southeast (16%),<sup>13</sup> Central West (11.5%),<sup>8</sup> Northeast (5%),<sup>14</sup> and

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

<sup>2</sup>Instituto Gonçalo Moniz, Fundação Oswaldo Cruz—FIOCRUZ, Salvador, Brazil.

<sup>3</sup>Universidade Católica do Salvador, Salvador, Brazil.

<sup>4</sup>Instituto Oswaldo Cruz, Fundação Oswaldo Cruz—FIOCRUZ, Rio de Janeiro, Brazil.

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

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

North (2.9%)<sup>15</sup> 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.<sup>16–18</sup>

The Northeast Brazil consists of nine states and is the third region with the highest number of HIV-1 infections in the country.<sup>19</sup> 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,<sup>7,14,20,21</sup> 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](http://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-

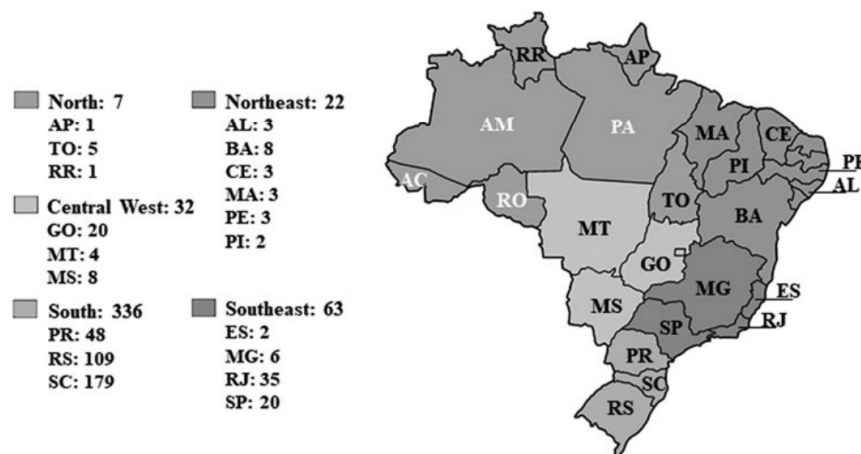
sponding to positions 2301–3200 relative to HXB2 reference strain. To compose the Brazilian dataset, a total of 948 subtype C sequences covering the genomic fragment aforementioned were initially collected. Subsequently, sequences were excluded according to the following criteria: absence of information 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 tool<sup>22</sup> 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 software<sup>23</sup> and submitted to HIV BLAST online tool against all subtype C sequences ([https://www.hiv.lanl.gov/content/sequence/BASIC\\_BLAST/basic\\_blast.html](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 program<sup>24</sup> (<http://mafft.cbrc.jp>) under the command: `mafft —thread 8 —threadtb 5 —threadit 0 —reorder —auto input > output` and manually edited using BioEdit software.<sup>25</sup> 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.<sup>26</sup>

Maximum likelihood (ML) phylogenies were reconstructed using IQ-TREE 1.6.8 webserver.<sup>27</sup> The reconstruction was performed under the GTR+I+G nucleotide substitution model that was inferred in Modelfinder.<sup>28</sup> 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)<sup>29</sup> 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 GENEDOC<sup>30</sup> 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 software<sup>23</sup> 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.<sup>31</sup> The Bayesian analyses were conducted in BEAST 1.10.4 package.<sup>32–34</sup> The ML phylogeny was used as a starting tree for Bayesian time-scaled phylogenetic analysis using BEAST 1.10.4 in CIPRES SCIENCE GATEWAY.<sup>32,33,35</sup> The tree reconstruction was inferred with the GTR +  $\Gamma$ 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.<sup>36</sup>

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).<sup>18</sup> 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.<sup>37</sup> 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.<sup>35</sup>

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.<sup>38</sup>

## **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).

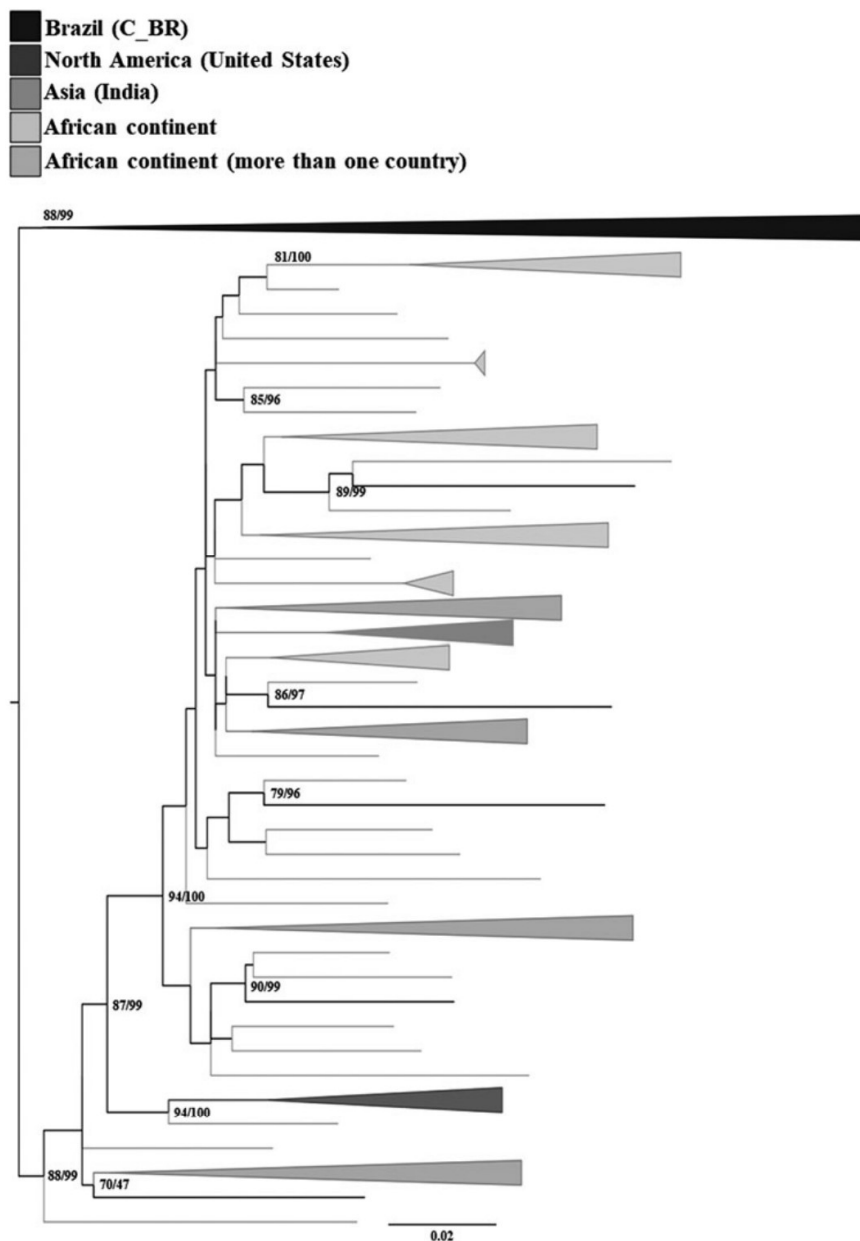
The presence of characteristic amino acid substitutions was investigated among the sequences representing the six putative subtype C introduction in Brazil (Supplementary Material S2). In the fragment of *pol* analyzed here, 14 motifs were conserved in all sequences. These predicted regions correspond 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.

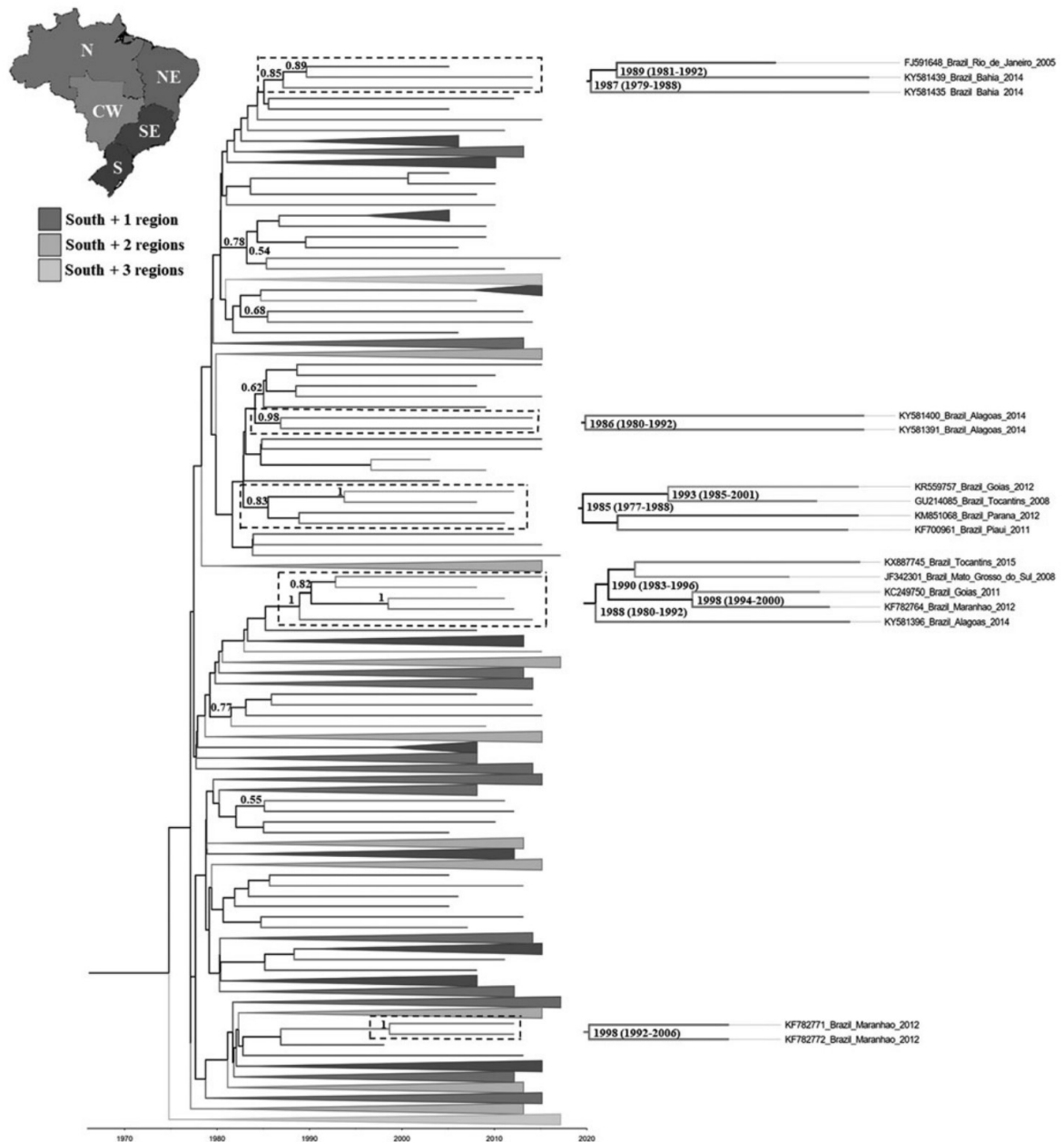
ML analysis showed that the 22 subtype C sequences from Northeast region appeared to be distributed among 19



**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  $2.068 \times 10^{-3}$  substitutions/sites/year (95% Bayesian Credibility Interval  $1.0-2.6 \times 10^{-3}$ ), consistent with other estimates for this subtype.<sup>17,18</sup> 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 (tMRCA)



**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.<sup>39</sup> 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.<sup>39</sup> In some countries, subtype C became the prevalent HIV-1 form after its introduction, surpassing another previously dominant genotype, such as in South Africa<sup>40,41</sup> and India.<sup>42,43</sup> 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.<sup>11,18</sup>

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.<sup>17,18</sup>

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.

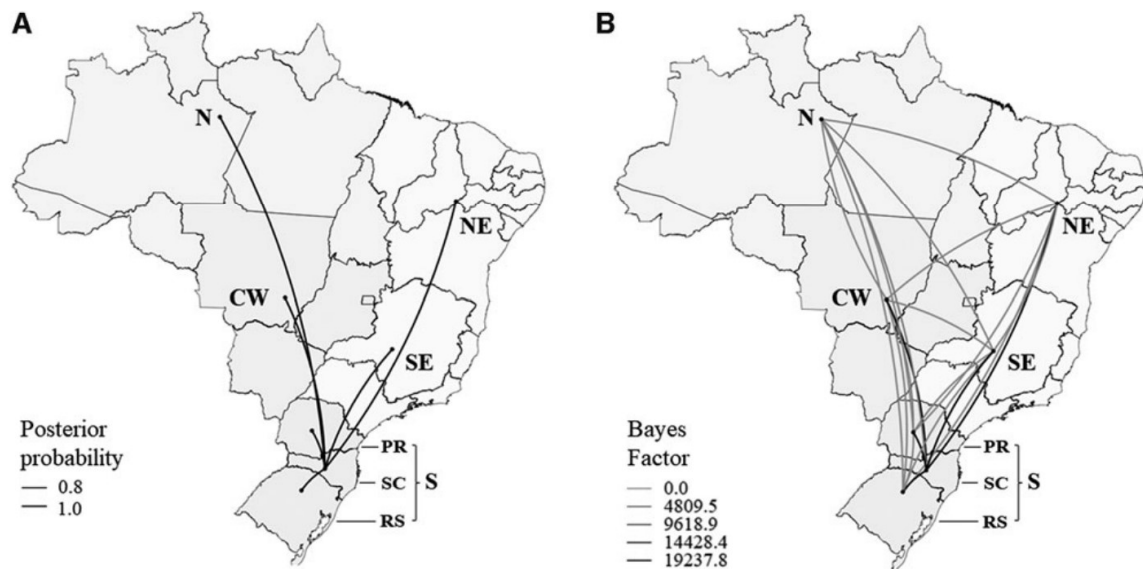


FIG. 4. Viral dispersion dynamics of HIV-1 C in Brazil. Phylogeographic analysis showing subtype C transmission route from South and it to all other Brazilian geographic regions. Viral strains from Northeast originated from South. The lines represent the supported transitions with a PP=1.0 (A) and Bayes 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.<sup>18</sup> Here, we bring evidence of an additional introduction of subtype C in Brazil represented by a strain isolated in Roraima<sup>44</sup> (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.<sup>45,46</sup> According to previous studies, the Brazilian subtype C indeed reveals specific amino acid patterns.<sup>47,48</sup>

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.<sup>47</sup> 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.<sup>48</sup> 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.<sup>14</sup> 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.<sup>18,49</sup> These studies indicated the South region as the main point of initial dispersion of the subtype C in Brazil, especially to Central West region,<sup>17,18,49</sup> which presents 7.8% of HIV-1 subtype C prevalence.<sup>9</sup> 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.<sup>50,51</sup> In fact, it has been shown that most of the subtype C strains found in Central West region present close phylogenetic re-

lationships with those isolated in the South.<sup>8,50</sup> Other studies reported that the Southeast region may also represent an additional strategic point for subtype C dispersion to other regions.<sup>8,51</sup> Delatorre and colleagues<sup>18</sup> 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.<sup>49</sup> 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 collaborators<sup>18</sup> 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.<sup>49</sup>

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.<sup>17,18</sup> 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.<sup>49</sup> 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,<sup>52</sup> attracting millions of visitors every year, 20% of which

are from South and Southeast regions.<sup>53</sup> 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

1. Ioachim HL: Immunopathogenesis of human immunodeficiency virus infection. *Cancer Res* 1990;50:5612–5617.
2. 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.
3. Santoro MM, Perno CF: HIV-1 genetic variability and clinical implications. *ISRN Microbiol* 2013;2013:481314.
4. Robertson DL, Anderson JP, Bradac JA: HIV-1 nomenclature proposal. *Science* 2000;288:55–56.
5. Silva LA, Divino F, Rêgo MO, *et al.*: HIV-1 Genetic diversity and transmitted drug resistance in antiretroviral treatment-naïve individuals from Amapá state, Northern Brazil. *AIDS Res Hum Retroviruses* 2016;32:373–376.
6. 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.
7. 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.
8. 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.
9. Arruda MB, Boullosa 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.
10. 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.
11. 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.
12. 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.
13. Alves BM, Siqueira JD, Prellwitz IM, *et al.*: Estimating HIV-1 genetic diversity in Brazil through next-generation sequencing. *Front Microbiol* 2019;10:749.
14. 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.
15. 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.
16. Bello G, Passaes CPB, Guimarães ML, *et al.*: Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS* 2008;22:1993–2000.
17. Bello G, Zanotto PMA, Iamarino A, *et al.*: Phylogeographic analysis of hiv-1 subtype c dissemination in Southern Brazil. *PLoS One* 2012;7:e35649.
18. 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.
19. 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](http://www.aids.gov.br/pt-br/pub/2019/boletim-epidemiologico-de-hivaids-2019) (2019), accessed June 11, 2020.
20. 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, Jermin 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: Spread3: 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, Chaturbui 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

## Capítulo III

---

Análise molecular e filogenética do HIV-1 subtipo C no estado da Bahia, Nordeste do Brasil.

(artigo: *Molecular and phylogenetic analysis of HIV-1 subtype C in Bahia state, northeastern Brazil*).

O subtipo C do HIV-1 está associado a mais da metade das infecções no sul do Brasil e vem aumentando em outras regiões do país. Em um estudo anterior realizado no nordeste do Brasil, encontramos uma prevalência de 4,1% do subtipo C. O presente trabalho investiga a origem do subtipo C no estado da Bahia com base em cinco novas sequências virais. A análise filogenética mostrou que os vírus do subtipo C encontrados na Bahia descendem da linhagem principal que circula em outras regiões brasileiras.

*O referido artigo foi submetido à revista AIDS Research and Human Retroviruses em 03/03/23.*

**Molecular and phylogenetic analysis of HIV-1 subtype C in Bahia state, northeastern Brazil.**

Running title: **Phylogenetic study of HIV-1 C in Bahia state, Brazil**

Rodrigo Cunha Oliveira<sup>a</sup>, Juliana Sacramento Mota de Souza<sup>a</sup>, Luiz Carlos Júnior Alcântara<sup>b</sup>,  
Monick Lindenmeyer Guimarães<sup>c</sup>, Carlos Brites<sup>d</sup> and Joana Paixão Monteiro Cunha<sup>a1</sup>

Author's affiliation:

<sup>a</sup>Núcleo de Bioinformática, Universidade Federal da Bahia, Departamento de Bioquímica e Biofísica, Salvador (Bahia, Brasil);

<sup>b</sup>Instituto Oswaldo Cruz Fundação Oswaldo Cruz, Laboratório de Flavivírus, Rio de Janeiro (Rio de Janeiro – Brazil);

<sup>c</sup>Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz - FIOCRUZ, Rio de Janeiro (Rio de Janeiro, Brazil);

<sup>d</sup>Complexo Hospitalar Prof. Edgard Santos, Laboratório de Pesquisa em Infectologia, Universidade Federal da Bahia, Salvador (Bahia, Brasil).

## ABSTRACT

HIV-1 Subtype C is associated with more than half of infections in southern Brazil and has been increasing in other regions of the country. In a previous study carried out in northeastern Brazil, we found a prevalence of 4.1% of subtype C. The present work investigate the origin of subtype C in the state of Bahia based on five new viral sequences. The phylogenetic analysis showed that subtype C viruses found in Bahia descend from the main lineage that circulates in other Brazilian regions.

HIV-1 infections in Brazil affect mainly young adults between 20-34 years (52.7%) and key populations, including men who have sex with men (MSM).<sup>1</sup> Northeastern Brazil ranks third in the highest number of HIV infections, comprising 19% of cases in the country, after the southeastern (44%) and southern regions (20%).<sup>1</sup> In HIV-1 Brazilian epidemic, the B, C, F1 subtypes and, BC and BF recombinants are the main circulating strains.<sup>2</sup> Due to the large territorial extension, the HIV-1 molecular epidemic presents unequal distribution among the five Brazilian major regions.<sup>3</sup> HIV-1 subtype C (HIV-1C) is the dominant genotype in the southern region (>50%), and the prevalence rates in the population decrease in the northernmost regions: <20% in the Central West and <10% in the Southeast, North and Northeast.<sup>2,3</sup> Although the genotypes B and BF are prevalent in the Northeast, the subtype C accounts for 5% of infections in this region<sup>4</sup> and some studies highlight that the increase in C strain over time is discreet.<sup>2,3</sup> This increase may be related to their biological properties, causing long periods without the onset of symptoms with slow disease progression leading to greater chances of transmission.<sup>2</sup>

The subtype C prevalence in Bahia state was 1.7% in 2011<sup>5</sup> and 5% in 2017<sup>4</sup>. Samples presented here were collected in 2019 from HIV-1 infected patients followed up at Professor Edgard Santos University Hospital in Salvador (Bahia/Brazil). The present study analyzed five new HIV-1 subtype C sequences (Table 1) from a population study (data not published) of the prevalence of HIV-1 subtypes in Bahia. These sequences were obtained as previously described.<sup>6</sup> One of them (011) is the first subtype C Near Full Length Genome (NFLG) sequenced in the state. For the other four samples, it was possible to amplify only partial genome fragments (Figure 1).

Two subtype C reference datasets was used to investigate the phylogenetic origin of the Bahia subtype C sequences, the first one containing 570 (460 from Brazil and 110 from world) sequences, based on the *pol* gene fragment (nucleotides 2301-3200 relative to HXB2 reference

strain)<sup>7</sup>, and a second dataset including 2,105 NFLG worldwide available in the Los Alamos database was created to analyze the subtype C NFLG sequence. Duplicated entries were removed and 1,915 sequences (Supplementary material 1) were used for phylogenetic reconstruction with the NFLG sample (011) obtained in the present study. Alignment was performed using the MAFFT online platform.<sup>8</sup> Maximum likelihood (ML) phylogenies were reconstructed using IQ-TREE.<sup>9</sup> HIVdb (<https://hivdb.stanford.edu>) was used to verify the HIV antiretroviral drug resistance.

Figure 2a shows the *pol* phylogenetic relationships of four new HIV-1 subtype C sequences from Bahia (007, 011, 136 and 224) plus reference sequences. All samples from Bahia grouped inside the largest Brazilian monophyletic cluster (Bootstrap [BS] = 98 and 99, respectively) which represents the main subtype C lineage that circulates in this country. Two sequences (007 and 011) clustered with samples from southern and southeastern Brazilian regions (states of Paraná, Rio de Janeiro and Santa Catarina) (BS 94 and 97 respectively). One sample (136) clustered with sequences from four Brazilian regions (South, Southeast, Central-West and Northeast), including one sequence from Pernambuco, another state from the Northeast region (BS 97). Another sample (224) clustered in a large group containing samples from three regions of Brazil: South, Southeast and Central-West (BS 92) (Figure 2a). Finally, sample 112 that could only be amplified in a different genomic region (positions 4171-5196 relative to HXB2). Thus, it was phylogenetic analyzed separately (Figure 2b) but also clustered inside the main Brazilian subtype C cluster (aLRT/BS = 84/81), again, close to a strain from Pernambuco.

In the NFLG dataset containing 1,916 (including sample 011 obtained in the present study) sequences, 26 sequences were described in Brazil, one in Argentina and one in Uruguay. The tree topology based on NFLG reveals that all South American C samples formed a monophyletic group (BS 100) separate from other subtype C sequences (supplementary material 2). Among the Brazilian subtype C sequences, the 011 was related to strains from the southern and southeastern regions, similar to what was found in the analysis of the *pol* region (Figure 2).

Four (007, 011, 136 and 224) of five patients were being treated with antiretroviral therapy and four major DRMs were found in three of them: D30N (protease), K103T, E138K/A (reverse transcriptase) and V151L (integrase) (Table 1).

In Brazil, HIV-1C was first detected in 1992 in the city of Porto Alegre, the capital of Rio Grande do Sul, a southern Brazilian state.<sup>10</sup> Subsequent phylogeographic analysis estimated that this strain has been circulating in the south region since the 1970s<sup>11</sup> and afterwards spread

throughout other four regions of the country: Central-West and Southeast in the early 1980s<sup>12</sup>, Northeast in the mid-1980s<sup>7</sup> and North at the end of the 1980s.<sup>13</sup> Despite of this, there is evidence of multiple introductions of HIV-1C in Brazil.<sup>7,11</sup> Viruses found in northernmost Brazilian states were phylogenetically linked to those found in the southern region (the main Brazilian lineage).<sup>7,11,12</sup> Our previous analyses have shown that the C subtype strains found in the Northeast region are descendant of this lineage.<sup>7</sup> Accordingly, despite the evidence of additional introductions of subtype C in Brazil, the new sequences obtained in the present study were also inside the main cluster, related to strains from the southern, southeastern and central-western regions (Figure 2).

The ML analysis based on the NFLG (Supplementary material) shows a specific cluster composed of sequences from South America and that sample 011 from Bahia is also related to viral sequences from Brazilian southern and southeastern regions, like in the *pol* ML tree (Figure 2).

In the DRM analysis, the D30N substitution was found in one sequence (224) conferring resistance to Nelfinavir, a Protease Inhibitor (PI). The D30N decreases the enzymatic activity of the mutated structure and consequently reduces the replicative capacity of the virus.<sup>13,14</sup> In addition, the substitution E138A found in 224 and 011 reduces the effectiveness of Etravirine and Rilpivirine, which are Non-nucleoside Reverse Transcriptase Inhibitor (NNRTI) drugs, as well as K103T in sample 136 (patient using Nevirapine). In the analysis of mutations associated with resistance to Integrase Inhibitors (INSTI), the V151L, found in one sequence (136), is a rare substitution that confers resistance to Raltegravir and Elvitegravir. In DRM studies, it is common to expect mutations that reduce the action of NNRTIs, especially in K103X, since this class represents the first regimen for initial treatment.<sup>14</sup>

In summary, this study analyzes five new HIV-1 subtype C viruses from Bahia state, which seems to descend from the main lineage of this subtype that reached the southern region of Brazil in the mid-70s.<sup>11</sup> One of them comprises the first NFLG of subtype C obtained in the state. Besides, in agreement with our research group previous result, two new subtype C sequences grouped with another Northeastern sample, providing new evidence of the intraregional spreading.

**Accession numbers:** OR015786, OR015787, OR015788, OR015789 and OR015790.

#### **CONFLICT OF INTEREST**

There is no conflict of interest among co-authors.

## FUNDING

R.C.O. and J.S.M.S., were supported by scholarships (FAPESB). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## ACKNOWLEDGEMENTS

R.C.O. and J.S.M.S are grateful to Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for the Phd. scholarship.

R.C.O., J.S.M.S. and J.P.M-C. are grateful to C.B. and M.L.G. for advanced support to the research.

## REFERENCES

1. Ministério da Saúde do Brasil (2021) Boletim Epidemiológico HIV / Aids | 2021
2. Gräf T, Bello G, Andrade P, et al. HIV-1 molecular diversity in Brazil unveiled by 10 years of sampling by the national genotyping network. *Scientific Reports* 2021, 11(1): 15842. doi:10.1038/s41598-021-94542-5
3. Arruda MB, Boulosa LT, Cardoso CC, et al. Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): a survey of treatment-naïve individuals. *Journal of the International AIDS Society* 2018, 21(3), e25032. <https://doi.org/doi:10.1002/jia2.25032>
4. Delatorre E, Couto-Fernandez JC, Bello G. HIV-1 Genetic Diversity in Northeastern Brazil: High Prevalence of Non-B Subtypes. *AIDS Research and Human Retroviruses* 2017, 33(7), 639–647. <https://doi.org/doi:10.1089/aid.2017.0045>
5. Monteiro-Cunha JP, Araujo AF, Santos E, et al. Lack of High-Level Resistance Mutations in HIV Type 1 BF Recombinant Strains Circulating in Northeast Brazil. *AIDS Research and Human Retroviruses* 2011, 27(6), 623–631. doi:10.1089/aid.2010.0126
6. Reis MNG, Guimarães ML, Bello G, et al. Identification of new HIV-1 circulating recombinant forms CRF81-cpx and CRF99-BF1 in central Western Brazil and of unique BF1 recombinant forms. *Front Microbiol*, 2019. 10:1–13. <https://doi.org/10.3389/fmicb.2019.00097>
7. Oliveira RC, Graf T, Rego FFA, et al. Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region. *AIDS Research and Human Retroviruses* 2021; 37 (12): 913-921. <https://doi.org/10.1089/aid.2020.0308>
8. Katoh K, Kuma K, Toh H. et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 2005; 33(2): 511–518. <https://doi.org/10.1093/nar/gki198>
9. Nguyen LT, Schmidt HA, Haeseler AV, et al. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2015; 32 (1): 268–274. <https://doi.org/10.1093/molbev/msu300>
10. Csillag C. HIV-1 subtype C in Brazil. *Lancet*, 1994. 344:1354. [https://doi.org/10.1016/S0140-6736\(94\)90703-X](https://doi.org/10.1016/S0140-6736(94)90703-X)
11. 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 (9): e74072. <https://doi.org/10.1371/journal.pone.0074072>.

12. Graf T, Vrancken B, Junqueira DM, et al. Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *Journal of Virology* 2015; 89 (24): 12341-12348. <https://doi.org/10.1128/JVI.01681-15>
13. Baxter JD, Dunn D, Tostevin A, et al. Transmitted HIV-1 drug resistance in a large international cohort using next-generation sequencing: results from the Strategic Timing of Antiretroviral Treatment (START) study. *HIV Med* 2020; 22 (5): 360–371. <https://doi.org/10.1111/hiv.13038>
14. Inocencio LA., Pereira AA, Sucupira MCA, et al. Brazilian Network for HIV Drug Resistance Surveillance: a survey of individuals recently diagnosed with HIV. *Journal of the International AIDS Society* 2009; 12: 20. <https://doi.org/10.1186/1758-2652-12-20>

Corresponding author:  
Joana P. Monteiro-Cunha

Núcleo de Bioinformática, Departamento de Bioquímica e Biofísica, Universidade Federal da Bahia. Av. Reitor Miguel Calmon, s/n, Salvador, Brasil# 55 71 3283-8914  
- Fax. # 55 71 3283-8894. Email: [joana.monteiro@ufba.br](mailto:joana.monteiro@ufba.br)

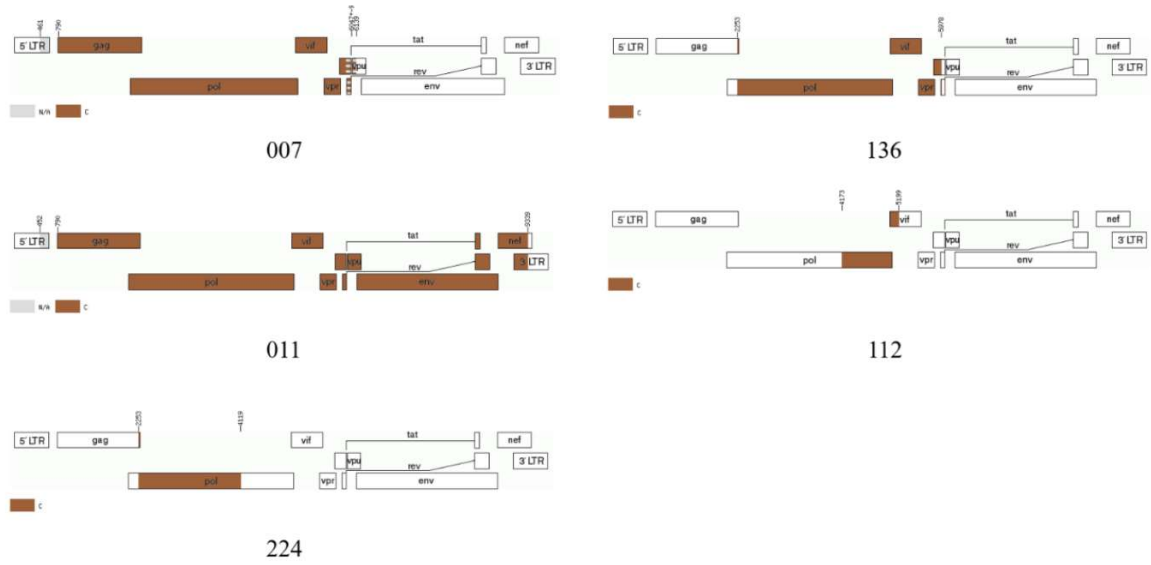


**Table 1:** Socio-demographic characteristics data from HIV-1 subtype C individuals from Bahia.

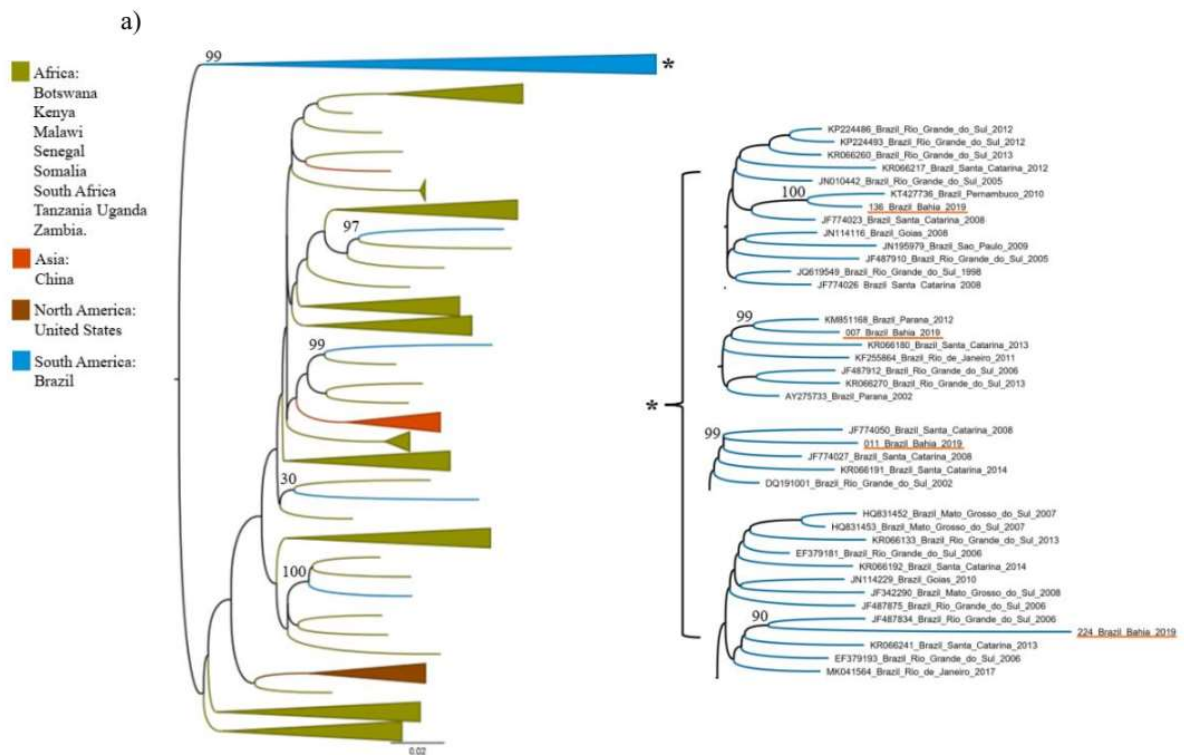
<i>Patient code</i>	<b>Gender</b>	<b>Diagnosis year</b>	<b>Origin</b>	<b>Subtype</b>	<b>DRM</b>	
					Protease/reverse transcriptase	Integrase
007	Female	2003	Cairú - Bahia	C	-	-
011	Male	2010	Salvador - Bahia	C	E138A	-
112	Male	2018	Salvador - Bahia	C	NA	-
136	Female	2009	Salvador - Bahia	C	K103T /	V151L
224	Male	2006	Salvador - Bahia	C	D30N / E138K	NA

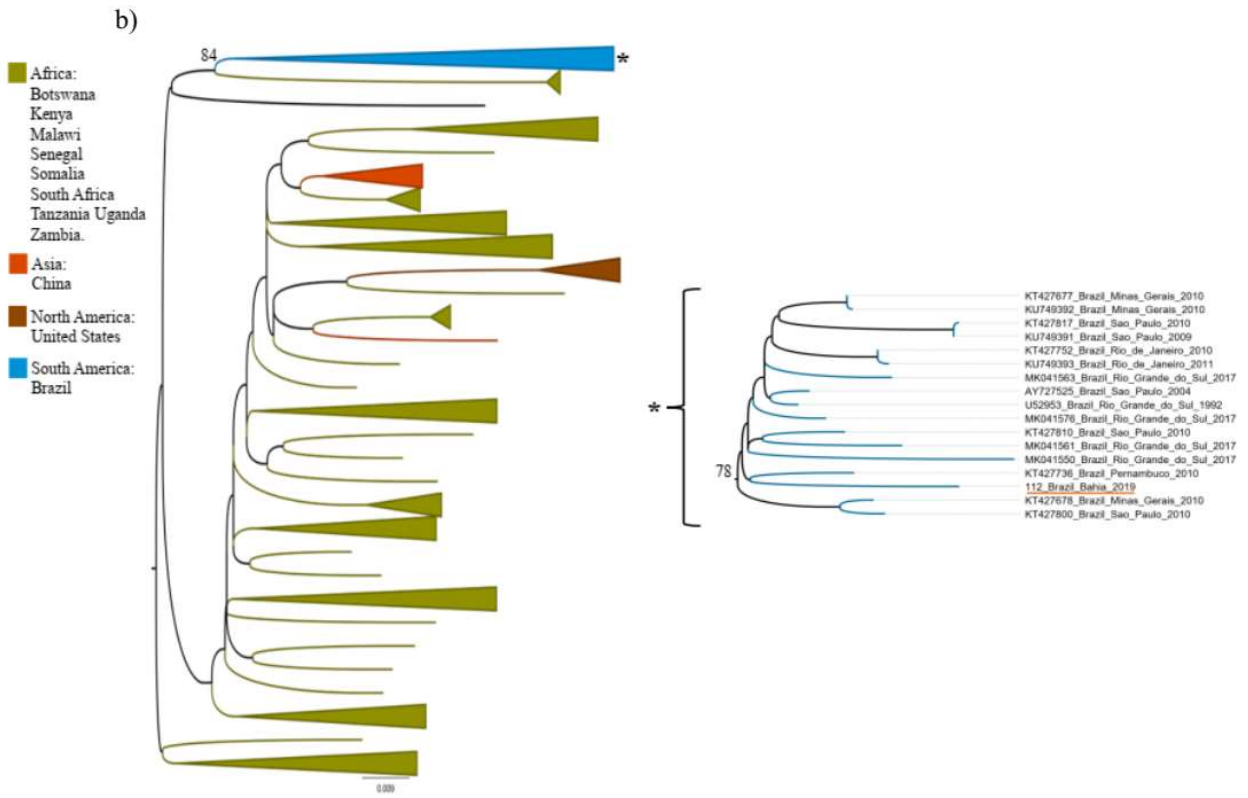
DRM- drug resistance mutations, NA- not amplified, A - Alanine, D – Aspartic acid, E – Glutamic acid, K - Lisine, N - Asparagine, T – Threonine, V - Valine.

**Figure 1** - Sequenced genomic regions of HIV-1 subtype C viral isolates found in Bahia state, Brazil.



**Figure 2** - Maximum likelihood analyses showing phylogenetic relationships of HIV-1 subtype C sequences from Bahia with sequences from Brazil and worldwide. The trees were rooted by the midpoint and the color of the branches represents the geographic origin of sequences. The statistical support is indicated only at key nodes as bootstrap values. Trees were built under the evolutionary model GTR+I+G+F and visualized in the Figtree v1.4.4 software. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating 0.02 nucleotide substitutions per site. **(a)** Samples 007, 011, 136 and 224 plus 570 HIV-1C reference sequences, based on a *pol* fragment (nucleotides 2301-3200 relative to HXB2 reference strain). Additional introductions of subtype C in Brazil are represented by the blue branches outside the main Brazilian cluster (\*) (collapsed). **(b)** Sample 112 plus 136 HIV-1C reference sequences, based on an integrase fragment (positions 4171-5196 relative to HXB2).

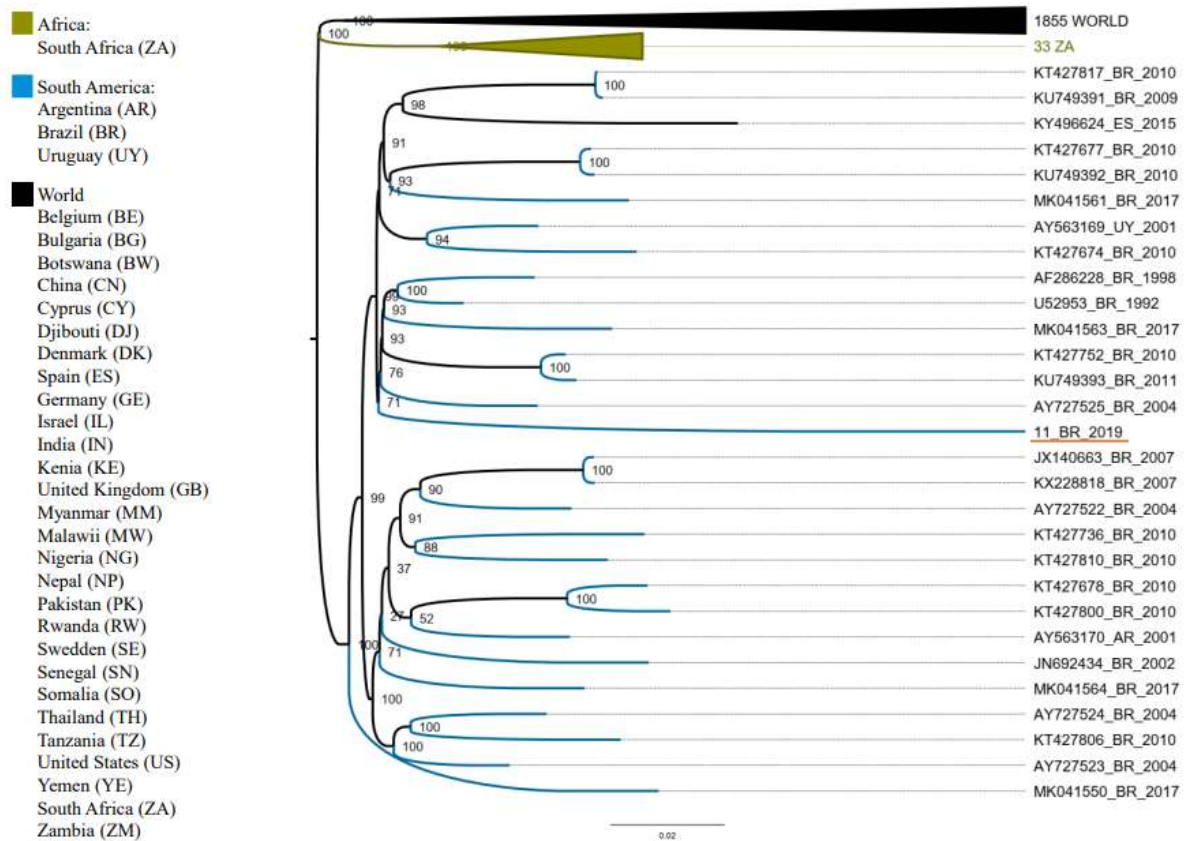




**Supplementary material 1-** Nucleotide sequence list  
<https://data.mendeley.com/datasets/p3nwyjsm6k/1>

**Supplementary material 2-** Maximum likelihood analyses showing phylogenetic relationships of the HIV-1 subtype C NFLG sequence from Bahia with subtype C sequences from Brazil (25) and worldwide (1980) (positions 838-7944 relative to HXB2). The tree was rooted by the midpoint and the color of the branches represents the geographic origin of sequences. The statistical support is indicated only at key nodes as bootstrap values. Trees were built under the evolutionary model GTR+I+G+F and visualized in the Figtree v1.4.4 software. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating 0.02 nucleotide substitutions per site.

<https://data.mendeley.com/datasets/tgcwvwfnp9/1>



## Capítulo IV

---

Uma nova Forma Recombinante Circulante (CRF) do HIV-1

B/C detectada em três regiões brasileiras.

(artigo: *A new HIV-1 B/C Circulating Recombinant Form (CRF) detected in three Brazilian regions*).

Três sequências nucleotídicas do HIV-1 pertencentes ao genótipo recombinante BC brasileiras foram encontradas em banco de dados públicos. Uma amostra parcial BC foi identificada em uma pessoa vivendo com HIV/AIDS acompanhado em hospital público, no estado da Bahia, Brasil em 2019. As análises de recombinação revelam que as quatro amostras possuem o mesmo padrão de recombinação no gene *pol*, a nível da região genômica que codifica a Integrase. Os estudos de filogenia realizados com as quatro amostras indicam fortemente uma relação filogenética entre as cepas. Portanto, acreditamos que uma nova forma recombinante circulante do HIV-1 no país, e as quatro amostras relacionam três de regiões estados brasileiras: Sul, Sudeste e Nordeste.

*O referido trabalho será submetido para publicação.*

## **A new HIV-1 B/C Circulating Recombinant Form detected in three Brazilian regions.**

Rodrigo Cunha Oliveira<sup>a</sup>, Darren Martin<sup>b</sup>, Juliana Sacramento Mota de Souza<sup>a</sup>, Luiz Carlos Júnior Alcântara<sup>c</sup>, Monick Lindenmeyer Guimarães<sup>d</sup>, Carlos Brites<sup>e</sup> and Joana Paixão Monteiro-Cunha<sup>a1</sup>

Author's affiliation:

<sup>a</sup>Núcleo de Bioinformática, Universidade Federal da Bahia, Departamento de Bioquímica e Biofísica, Salvador (Bahia, Brasil);

<sup>b</sup>Department of Integrative Biomedical Sciences, Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Anzio Road Observatory, Cape Town 7549, South Africa;

<sup>c</sup>Instituto Oswaldo Cruz Fundação Oswaldo Cruz, Laboratório de Flavivírus, Rio de Janeiro (Rio de Janeiro, Brazil);

<sup>d</sup>Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz - FIOCRUZ, Rio de Janeiro (Rio de Janeiro, Brazil);

<sup>e</sup>Complexo Hospitalar Prof. Edgard Santos, Laboratório de Pesquisa em Infectologia, Universidade Federal da Bahia, Salvador (Bahia, Brasil).

*Dear Editor R. Keith Reeves and Lishomwa Ndhlovu,*

To date, more than 120 HIV-1 Circulating Recombinant Forms (CRF) have been described worldwide and are listed in Los Alamos National Laboratory (LANL) database (<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). Fourteen of these were classified as CRF\_BC, of which, 11 were identified in China.

In 2021, three new BC CRFs were identified, two in China (110 and 118) and one in Spain (108) (CAÑADA et al., 2021; GAO et al., 2021; LI et al., 2021). In Brazil, 81% of the HIV-1 infections are caused by B and C subtypes viruses and a BC CRF (CRF31\_BC) has been identified in this country (GRAF et al., 2021; SANTOS et al., 2006). Furthermore, BC Unique Recombinant Forms (URF) are commonly detected in all Brazilian territory. Here we present a new recombinant genotype of HIV-1, indicating a new Brazilian CRF\_BC.

The sample presented here (223) were collected in 2019 from a HIV-1 infected patient followed

at Professor Edgard Santos University Hospital in Salvador (Bahia/Brazil). The study was approved by both Hospital and the Federal University of Bahia Ethics Committee. The participant supplied written informed consent for specimen collection and subsequent analyses. The PCR and DNA sequencing protocol were applied as described in (REIS et al., 2019). A genomic fragment corresponding to *integrase/vif* and *nef* genes (positions 4172 to 5198 and 8694 to 9512 relative to HXB2) (Figure 1) were sequenced through Sanger method, as previously described (REIS et al., 2019). The other HIV-1 genomic regions did not amplify successfully. RDP5 (MARTIN et al., 2021), JphMM (SCHULTZ et al., 2009) and Simplot (SALMINEN et al., 1996) recombination tools were used to evaluate the mosaic pattern. A breakpoint between C and B subtypes were found in the *integrase/vif* region (positions 4837+77 – 5071+21 HXB2 coordinates), while the *nef* gene region reveals a pure C fragment (Figure 1 and Supplementary material 1).

Using the above recombination tools, three other Brazilian former URFs, obtained from LANL database, were identified with similar breakpoints in *integrase/vif* and pure C in *nef* genes: GQ365652 (463 - 9632 of HXB2 strain), KT427778 (530 - 9537 of HXB2 strain) and KT427790 (521 - 9582 of HXB2 strain). According to the publications that originated these sequences (PASSAES et al., 2009; PESSOA et al., 2016), there was no epidemiological link between the infected individuals.

Comparing our BC recombinant fragment with the other three Near-Full Length Genome (NFLG) BC sequences from Brazil (Figure 1), we identified similar breakpoints: GQ365652 (4834+75 - 5085+21) from South region, KT427778 (4831+80 - 5080+19) and KT427790 (4851+75 - 5085+21) from Southeast region. After that, 137 nucleotide sequences belonging to the 14 BC CRFs patterns, plus 39 pure subtype reference sequences were retrieved from LANL database, aiming the Maximum Likelihood (ML) analyzes that were conducted through IQtree webserver (TRIFINOPOULOS et al., 2016) and RDP5 (MARTIN et al., 2021).

A total of 180 samples (Supplementary material 2) were used in this study and based in the sequence and the recombination pattern of sample 223, four ML trees were reconstructed: figure 2A: BC *integrase* and C *nef* concatenated fragments; 2B: subtype C fragment (*integrase* and *nef*) concatenated; 2C: subtype B *integrase* fragment and 2D: NFLG sequences without the 223 sample. In the four trees all sequences (223, GQ365652, KT427778 and KT427790) grouped together with supported aLRT (approximated Likelihood Ratio Test) and Bootstrap (BS) values. In the ML trees obtained in RDP5 software reveals a similar topology (Supplementary material 3 and 4). Analyzing the similarity degree, the 223 sample showed high identity with



KT427778, GQ365652 and KT427790 (Supplementary material 4).

In Brazil, the subtype C was identified in the South region (CSILLAG, 1994), likewise the CRF31\_BC (SANTOS et al., 2006). This region accounts for 49,3%% of HIV-1 subtype C infections in Brazil (GRAF et al., 2021). Out of the four samples representing the new CRF, one (GQ365652) were described in Rio Grande do Sul (South region), two (KT427778 and KT427790) were from Rio de Janeiro state (Southeast region) and one (223) from Bahia state (Northeast region). In addition, Passaes and collaborators, (2009), found the same breakpoint (4834+-75 – 5085+-21) in URF\_BC sequences from Rio Grande do Sul state (South region) and suggested that second-generation recombinant events, involving BC recombinant strains as parental, may be favoring the increase of the new genotypes, once both pure subtypes are prevalent in the country.

In brief, here we purpose a CRF\_BC in Brazil, which are circulating in, at least, three out of five Brazilian geographic regions. The emergence of new genotypes maintains the red-light turn-on to the necessity of continuous molecular research of HIV-1. More specific studies may be necessary to understand the evolution mechanism and their pathogenesis.

## **CONFLICT OF INTEREST**

There is no conflict of interest among co-authors.

## **FUNDING**

R.C.O. was supported by scholarship. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## **ACKNOWLEDGEMENTS**

R.C.O. and J.S.M.S are grateful to Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for the Phd. scholarship.

R.C.O., J.S.M.S. and J.P.M. are grateful to D.M., C.B., L.C.J.A. and M.L.G. for advanced support to the research.

## **REFERENCES**

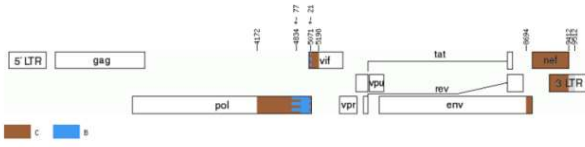
1. Cañada, J.E., Delgado, E., Gil, H., Sánchez, M., Benito, S., García-Bodas, E., Gómez-González, C., Canut-Blasco, A., Portu-Zapirain, J., de Adana, E.S., De la Peña, M., Ibarra, S., Cilla, G., Iribarren, J.A., Martínez-Sapiña, A., Thomson, M.M. Identification of a New

- HIV-1 BC Intersubtype Circulating Recombinant Form (CRF108\_BC) in Spain. *Viruses*. 2021; 13 (1): 93. <https://doi.org/10.3390/v13010093>
2. Gao, Q., Feng, Y., Gao, L., Zhang, M., Liu, Y., Xiao, M., Dong, X., Xia, X. Genetic characteristics of a New HIV-1 subtype B/C intersubtype circulating recombinant form (CRF118\_BC) identified in Yunnan, China. *Journal of Infection*. 2021; 83 (2): 237-279. <https://doi.org/10.1016/j.jinf.2021.05.006>
  3. Li, J., Gao, Q., Zhang, M., Liu, J., Jia, Y., Feng, Y., Xia, X., Dong, X. A newly emerging HIV-1 circulating recombinant form (CRF110\_BC) comprising subtype B and C among intravenous drug users in Yunnan, China. *Journal of Infection*. 2021; 82 (3): e8–e10. <https://doi.org/10.1016/j.jinf.2020.12.015>
  4. Gräf, T., Bello, G., Andrade, P., Arantes, I., Pereira, J. M., da Silva, A. B. P., Veiga, R. V., Mariani, D., Boullosa L, T., Arruda, M, B., Fernandez, J, C, C., Dennis, A, M., Rasmussen D, A., Tanuri, A. (2021). HIV-1 molecular diversity in Brazil unveiled by 10 years of sampling by the national genotyping network. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-94542->
  5. Santos, A.F., Sousa, T.M., Soares, E.A.J.M., Sanabani, S., Martinez, A.M.B., Sprinz, E., Silveira, J., Sabino, E.C., Tanuri, A., Soares, M.A. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS*. 2006; 20; (16): 2011-2019. <https://doi.org/10.1097/01.aids.0000247573.95880.db>.
  6. Reis MNG, Guimarães ML, Bello G, et al. Identification of new HIV-1 circulating recombinant forms CRF81-cpx and CRF99-BF1 in central Western Brazil and of unique BF1 recombinant forms. *Front Microbiol*, 2019. 10:1–13. <https://doi.org/10.3389/fmicb.2019.00097>
  7. Martin, D. P., Varsani, A., Roumagnac, P., Botha, G., Suresh, M., Schwab, Kelz, Z., Kumar, V., Murrel, B. RDP5: a computer program for analyzing recombination in, and

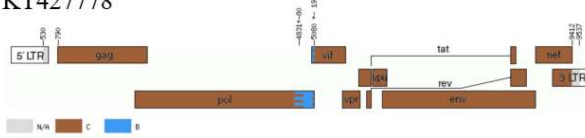
- removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*. 2021; 7 (1). <https://doi.org/10.1093/ve/veaa087>
8. Schultz, A.K., Zang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., Stanke, M. jpHMM: Improving the reability of recombination prediction in HIV-1. *Nucleic Acids Research*. 2009; 37 (Web Server): W647–W651. <https://doi.org/10.1093/nar/gkp371>
  9. Salminen, M.O., Johansson, B., Sonnerborg, A., Ayehunie, S., Gotte, D., Leinikki, P., Burke, D. S., Mccutchan, E. Full-length sequence of na Ethiopian Human Immunodeficiency Virus type-1 (HIV-1) isolate of genetic subtype C. *AIDS Research and Human Retroviruses*. 1996; 12, nº. 14, 1329-1339.
  10. Passaes, C.P.B., Guimarães, M.L., Bello, G., Morgado, M.G. Near Full-Length Genome Characterization of HIV Type 1 Unique BC Recombinant Forms from Southern Brazil. *Aids Research and Human Retroviruses*. 2009; 25: 12. <https://doi.org/10.1089=aid.2009.0167>
  11. Pessôa, R., Loureiro, P., Lopes, M.E., Carneiro-Proietti, A. B. F., Sabino, E. C., Busch, M. P., Sanabani, S. S., Ultra-Deep Sequencing of HIV-1 near Full-Length and Partial Proviral Genomes Reveals High Genetic Diversity among Brazilian Blood Donors. *PLoS One*. 2016; 11(3): e0152499. <https://doi.org/10.1371/journal.pone.0152499>
  12. Trifinopoulos, J., Nguyen, L.T., Haeseler, A.V., Minh, B.Q. W-IQ-TREE: a fast-online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*. 2016; 44 (W1): W232–W235. <https://doi.org/10.1093/nar/gkw256>
  13. Csillag C. HIV-1 subtype C in Brazil. *Lancet* 1994; 344:1354

**Figure 1-** Recombination patterns of HIV-1 BC strains.

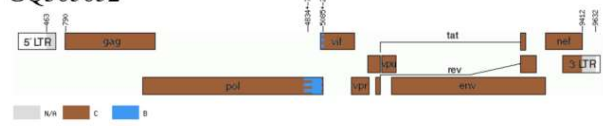
223



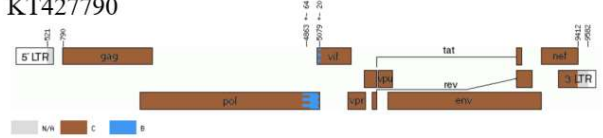
KT427778



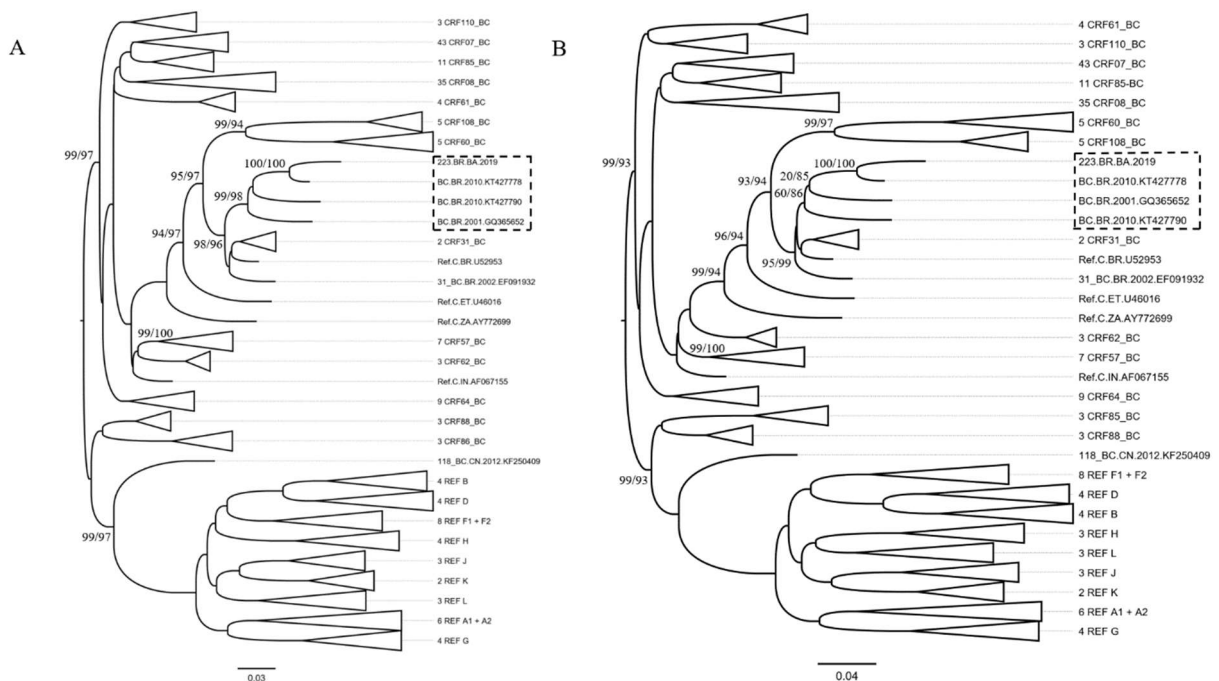
GQ365652

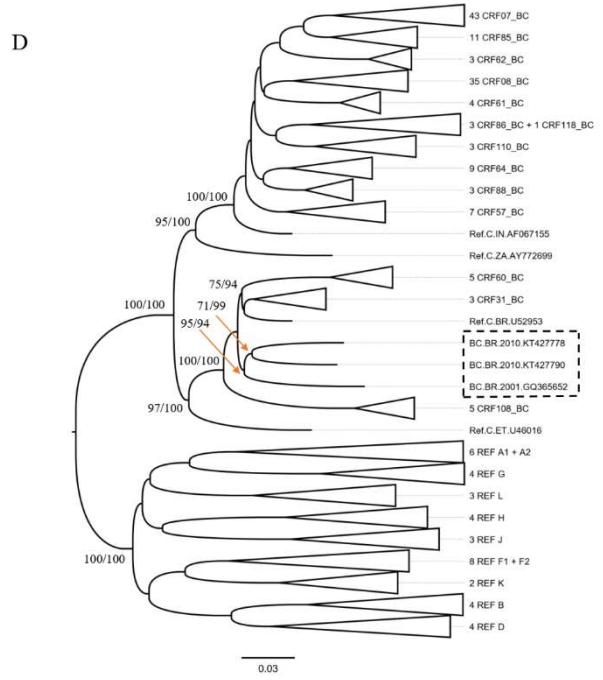
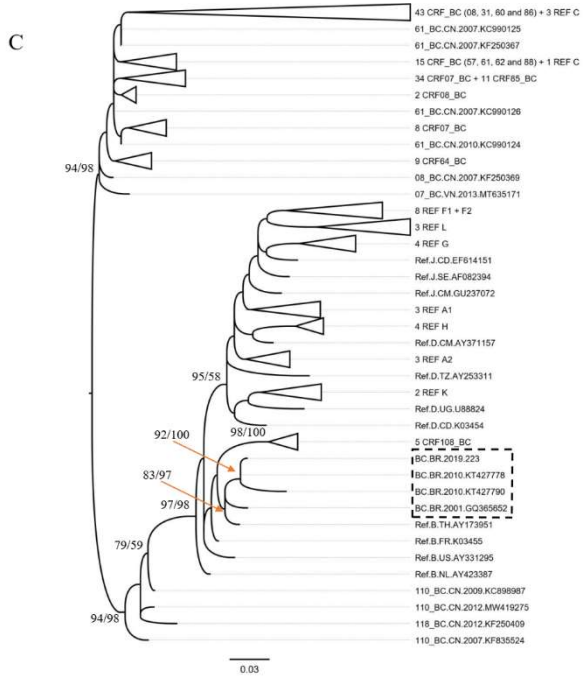


KT427790



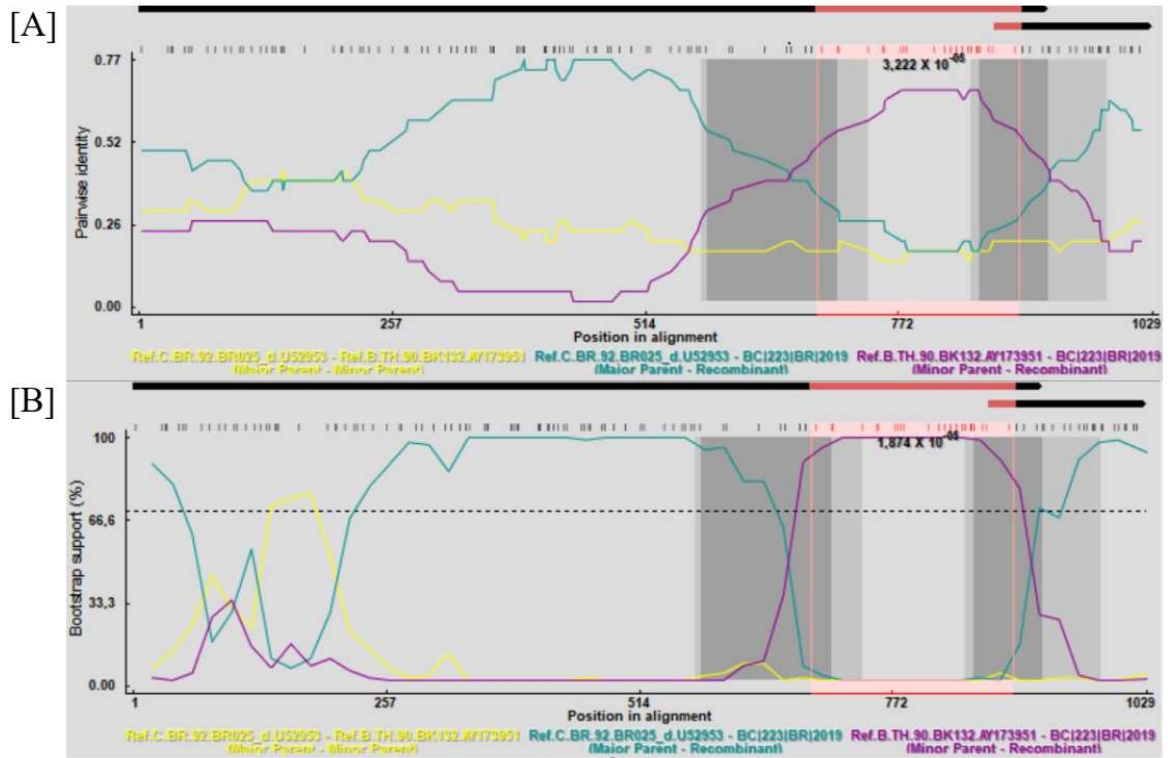
**Figure 2-** Maximum likelihood analyses showing phylogenetic relationship (**A, B, C and D**) of 180 HIV-1 sequences plus four recombination BC sequences from Bahia (223) and Brazil (GQ365652, KT42778 and KT427790). Of these 137 sequences belonging to 14 CRFs and 39 subtype reference sequences (the **(D)** tree contains 179 samples). The trees were rooted by the midpoint and in the collapsed clusters, the subtype classification (CRF) and number of sequences indicated. The statistical support is indicated only at key nodes as aLRT and bootstrap values. The trees (**A, B, C and D**) were built under the evolutionary model GTR+I+G+F and the **(C)** under TN+I+G+F model. The trees were visualized in the Figtree v1.4.4 software. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating the nucleotide substitutions per site. **(A)** BC *integrase* and C *nef* concatenated fragments; **(B)** subtype C fragment (*integrase* and *nef*) concatenated **(C)** B subtype *integrase* fragment; (nucleotides 4833 - 5066 relative to HXB2 reference strain) and **(D)** NFLG sequences without 223 (nucleotides 805 - 8912 relative to HXB2 reference strain).





**Supplementary material 1-** Recombination pattern detected of the integrase fragment from the 223 sample by RDP5 program through RDP (A) and Bootscan (B) methods.

<https://data.mendeley.com/datasets/dggs6dmzyb/1>



---

**Supplementary material 2- Nucleotide sequence list**


---

Subtype classification	Origin	Year	Access number
07_BC	CN	1997	AF286226
07_BC	CN	1998	AF286230
07_BC	CN	2002	AF503396
07_BC	CN	2005	EF368370
07_BC	CN	2005	EF368371
07_BC	CN	2005	EF368372
07_BC	CN	2005	EF420986
07_BC	CN	2005	KC503852
07_BC	CN	2005	KC503853
07_BC	CN	2005	KC503855
07_BC	CN	2006	JX392378
07_BC	CN	2006	JX392379
07_BC	CN	2006	JX392380
07_BC	CN	2006	JX392381
07_BC	CN	2006	JX392382
07_BC	CN	2006	JX392383
07_BC	CN	2006	JX392384
07_BC	CN	2006	KF250376
07_BC	CN	2007	JQ423923
07_BC	CN	2007	KC492737
07_BC	CN	2007	KC492738
07_BC	CN	2007	KF250366
07_BC	CN	2007	KF250371
07_BC	CN	2007	KF250372
07_BC	CN	2007	KF250373
07_BC	CN	2007	KF250374
07_BC	CN	2007	KF250375
07_BC	CN	2007	KF250377
07_BC	CN	2007	KF250385
07_BC	CN	2007	KF835494
07_BC	CN	2007	KF835510
07_BC	CN	2007	KF835511
07_BC	CN	2007	KF835532
07_BC	CN	2008	HQ215552
07_BC	CN	2008	JX960601
07_BC	CN	2009	JX960600
07_BC	CN	2009	JX960602
07_BC	CN	2009	KC503854
07_BC	GB	2014	MF109555
07_BC	TW	2004	DQ230841
07_BC	TW	2013	KF234628
07_BC	VN	2013	MT635171
08_BC	CN	1997	AY008715
08_BC	CN	1997	AY008716

---



08_BC	CN	1997	AY008717
08_BC	CN	1998	AF286229
08_BC	CN	2000	AB746344
08_BC	CN	2000	AB746345
08_BC	CN	2000	AB773884
08_BC	CN	2000	AB773885
08_BC	CN	2000	KC914396
08_BC	CN	2001	AB746342
08_BC	CN	2001	AB746343
08_BC	CN	2006	HM067748
08_BC	CN	2007	JF719819
08_BC	CN	2007	KF250368
08_BC	CN	2007	KF250369
08_BC	CN	2007	KF312642
08_BC	CN	2007	KF835493
08_BC	CN	2007	KF835495
08_BC	CN	2007	KF835496
08_BC	CN	2007	KF835497
08_BC	CN	2007	KF835500
08_BC	CN	2007	KF835506
08_BC	CN	2007	KF835507
08_BC	CN	2007	KF835508
08_BC	CN	2007	KF835509
08_BC	CN	2007	KF835512
08_BC	CN	2007	KF835529
08_BC	CN	2007	KF835530
08_BC	CN	2007	KF835531
08_BC	CN	2007	KF835534
08_BC	CN	2007	KF835535
08_BC	CN	2007	KF835541
08_BC	CN	2007	KF835545
08_BC	CN	2007	KF835547
08_BC	CN	2014	KF835501
108_BC	ES	2016	MN172222
108_BC	ES	2017	MN172225
108_BC	ES	2017	MT559132
108_BC	ES	2018	MN172223
108_BC	ES	2018	MN172224
110_BC	CN	2007	KF835524
110_BC	CN	2009	KC898987
110_BC	CN	2012	MW419275
118_BC	CN	2012	KF250409
223	BR	BA	2019
31_BC	BR	2002	EF091932
31_BC	BR	2004	AY727526
31_BC	BR	2004	AY727527
57_BC	CN	2007	HM776939

57_BC	CN	2009	JX679207
57_BC	CN	2009	KC870044
57_BC	CN	2009	KC898985
57_BC	CN	2009	KC898991
57_BC	CN	2009	KC899008
57_BC	CN	2010	KF250400
60_BC	BR	2002	JN692442
60_BC	FR	2006	EU448296
60_BC	IT	2011	KC899079
60_BC	IT	2011	KC899080
60_BC	IT	2011	KC899081
61_BC	CN	2007	KC990125
61_BC	CN	2007	KC990126
61_BC	CN	2007	KF250367
61_BC	CN	2010	KC990124
62_BC	CN	2010	KC870034
62_BC	CN	2010	KC870035
62_BC	CN	2010	KC870037
64_BC	CN	2009	KC870031
64_BC	CN	2009	KC870042
64_BC	CN	2009	KC870043
64_BC	CN	2009	KC898994
64_BC	CN	2009	KC899009
64_BC	CN	2010	KC870032
64_BC	CN	2010	KC870033
64_BC	CN	2010	KC870036
64_BC	CN	2010	KC870040
85_BC	CN	2012	MK258684
85_BC	CN	2014	KU992928
85_BC	CN	2014	KU992929
85_BC	CN	2014	KU992930
85_BC	CN	2014	KU992931
85_BC	CN	2014	KU992932
85_BC	CN	2014	KU992934
85_BC	CN	2014	KU992935
85_BC	CN	2014	KU992936
85_BC	CN	2014	KU992937
85_BC	CN	2014	MK258683
86_BC	CN	2013	KX582249
86_BC	CN	2013	KX582250
86_BC	CN	2013	KX582251
88_BC	CN	2005	KC898975
88_BC	CN	2005	KC898979
88_BC	CN	2009	KF250402
BC	BR	2001	GQ365652
BC	BR	2010	KT427778
BC	BR	2010	KT427790

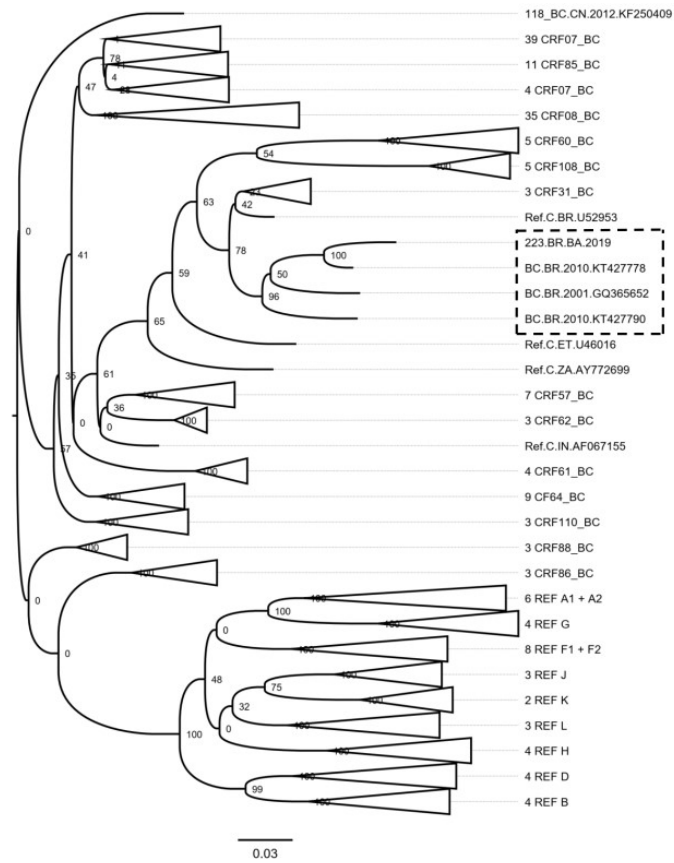
A1	AU	-	DQ676872
A1	RW	-	AB253421
A1	UG	-	AB253429
A2	CD	-	AF286238
A2	CM	-	GU201516
A2	CY	-	AF286237
B	FR	-	K03455
B	NL	-	AY423387
B	TH	-	AY173951
B	US	-	AY331295
C	BR	-	U52953
C	ET	-	U46016
C	IN	-	AF067155
C	ZA	-	AY772699
D	CD	-	K03454
D	CM	-	AY371157
D	TZ	-	AY253311
D	UG	-	U88824
F1	BE	-	AF077336
F1	BR	-	AF005494
F1	FI	-	AF075703
F1	FR	-	AJ249238
F2	CM	-	AF377956
F2	CM	-	AJ249236
F2	CM	-	AJ249237
F2	CM	-	AY371158
G	BE	-	AF084936
G	KE	-	AF061641
G	NG	-	U88826
G	PT	-	AY612637
H	BE	-	AF190127
H	BE	-	AF190128
H	CF	-	AF005496
H	GB	-	FJ711703
J	CD	-	EF614151
J	CM	-	GU237072
J	SE	-	AF082394
K	CD	-	AJ249235
K	CM	-	AJ249239
L	CD	-	AF286236
L	CD	-	AF457101
L	CD	-	MN271384

Legend: AU- Australia, BA- Bosnia Herzegovina, BE- Belgium, BR- Brazil, CD- Democratic Republic of Congo, CF- Central Africa Republic, CM- Cameroon, CN- China, CY- Cyprus, ES- Spain, ET- Ethiopia, FI- Finland, FR- France, GB- United Kingdom, IN- India, IT- Italy, KE- Kenya, NG- Nigeria, NL- Netherlands, PT- Portugal, RW- Rwanda, SE- Senegal, TH- Thailand, TW- Taiwan, TZ- Tanzania, UG- Uganda, US- United States, VN- Vietnam, ZA- South Africa.

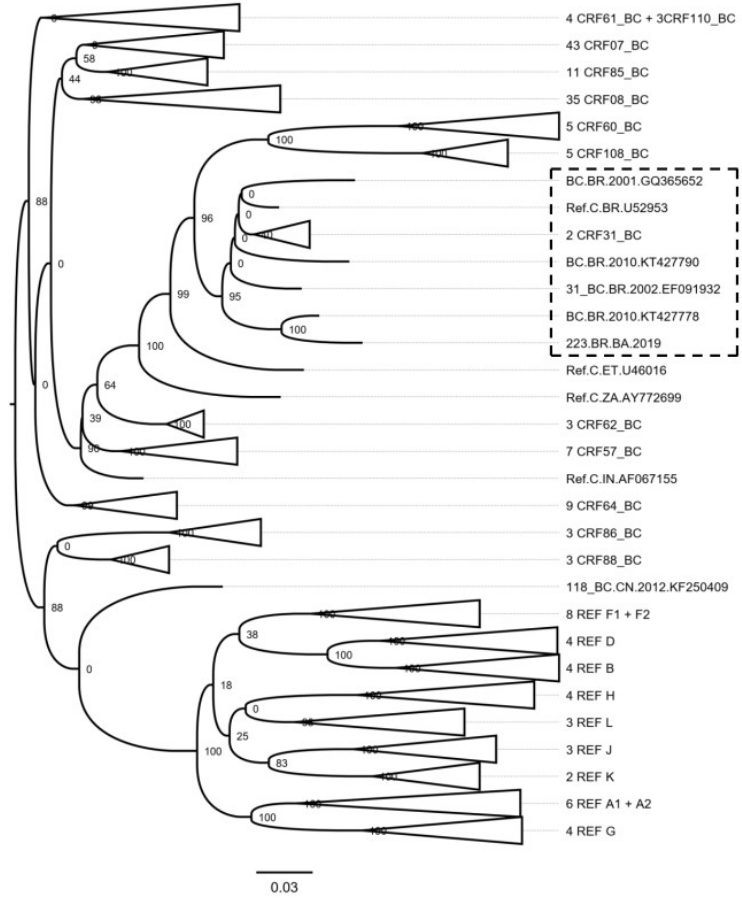
**Supplementary material 3-** RDP5 maximum likelihood analyses showing phylogenetic relationships (**A, B, C and D**) of 180 HIV-1 sequences plus four recombination BC sequences from Bahia (223) and Brazil (GQ365652, KT42778 and KT427790). Of these 137 sequences belonging to 14 CRFs and 39 subtype reference sequences (the **D**) tree contains 179 samples). The trees were rooted by the midpoint and in the collapsed clusters, the subtype classification (CRF) and number of sequences indicated. The statistical support is bootstrap values. The trees were visualized in the Figtree v1.4.4 software. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating the nucleotide substitutions per site. (**A**) BC *integrase* and C *nef* concatenated fragments; (**B**) subtype C fragment (*integrase* and *nef*) concatenated (**C**) B subtype *integrase* fragment; (nucleotides 4833 - 5066 relative to HXB2 reference strain) and (**D**) NFLG sequences without 223 (nucleotides 805 - 8912 relative to HXB2 reference strain).

<https://data.mendeley.com/datasets/dggs6dmzyb/1>

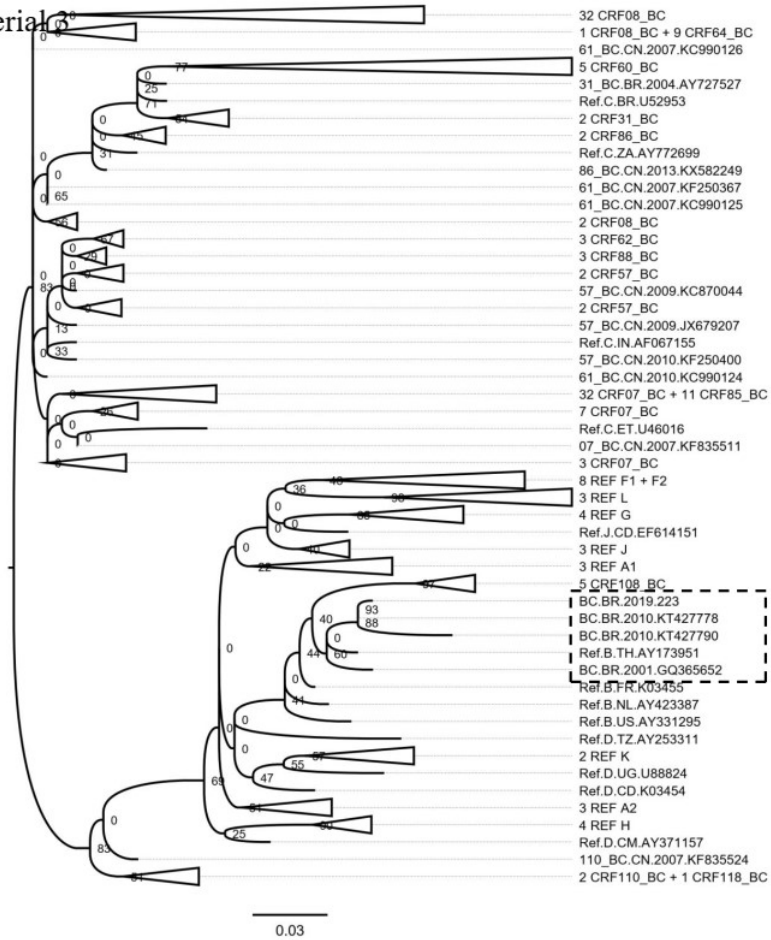
Supplementary material 3  
A



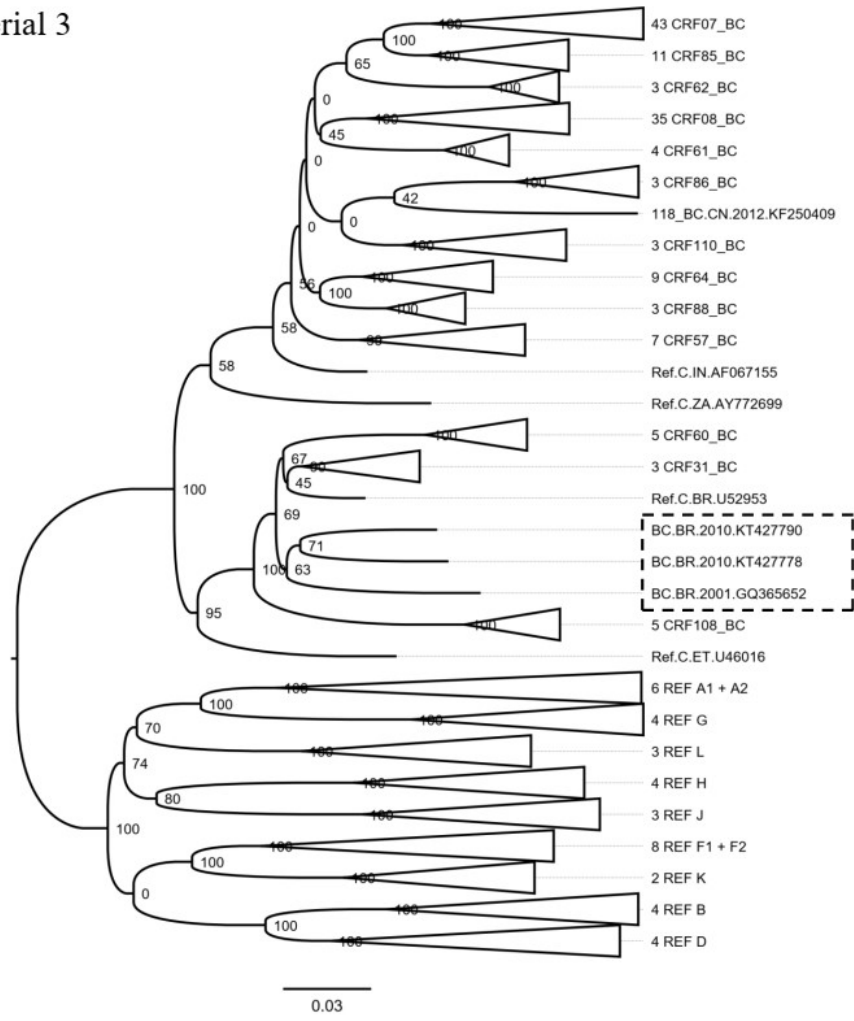
Supplementary material 3  
B



Supplementary material 3  
C



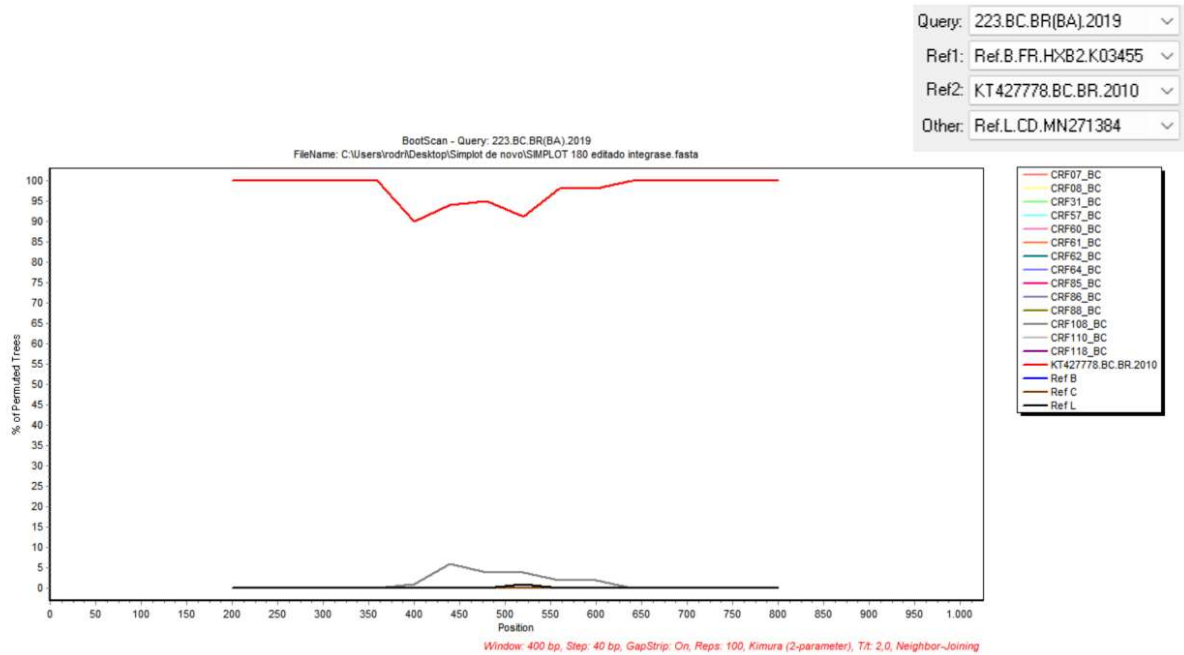
Supplementary material 3  
D



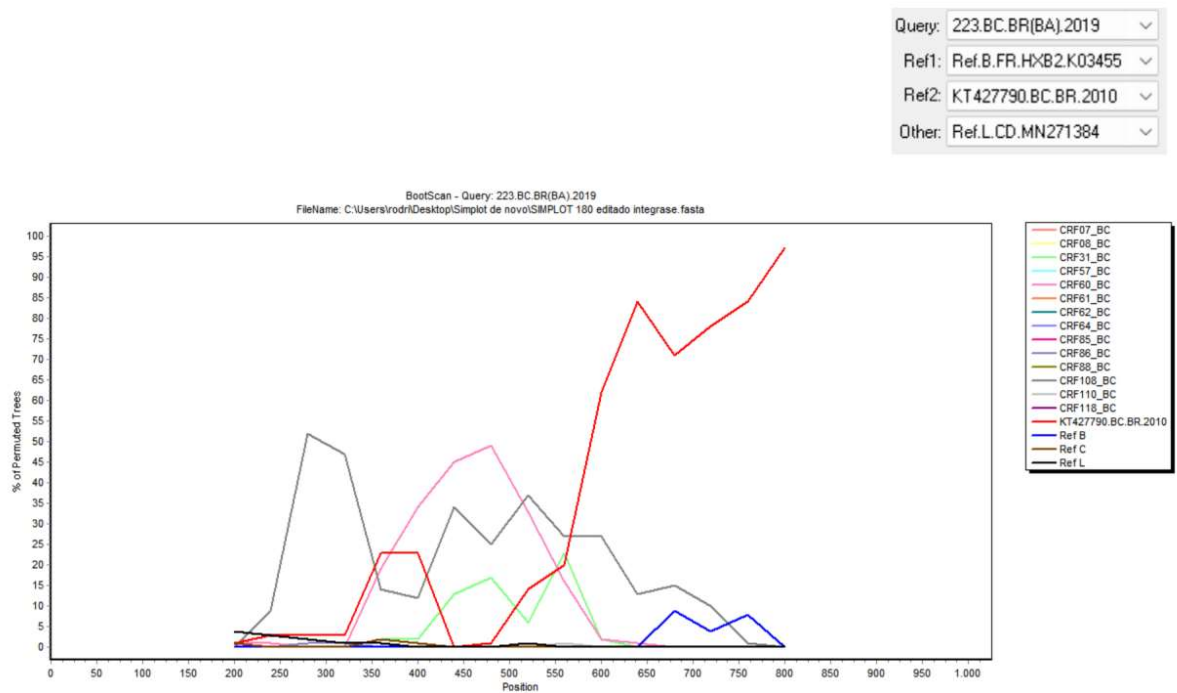
**Supplementary material 4-** Bootscan analysis (SimPlot) of the integrase fragment from the sample 223 against (A) KT427778, (B) KT427790 and (C) GQ365652. Sliding window of 400 base pairs moving in steps of 40 nucleotides.

<https://data.mendeley.com/datasets/dggs6dmzyb/1>

A

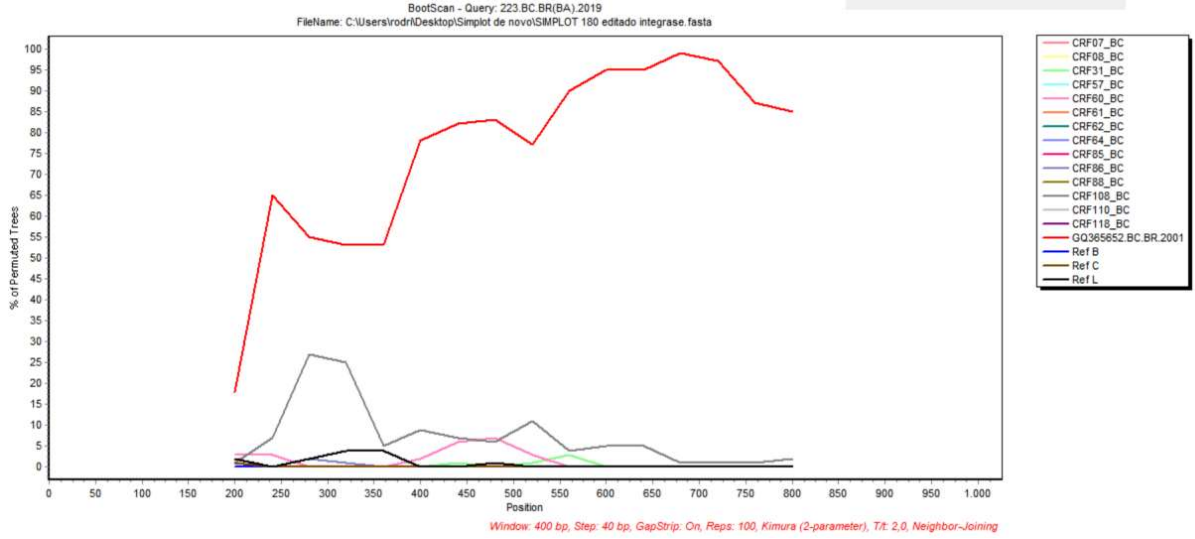


B



C

Query: 223.BC.BR(BA).2019  
Ref1: Ref.B.FR.HXB2.K03455  
Ref2: GQ365652.BC.BR.2001  
Other: Ref.L.CD.MN271384





## Capítulo V

---

### Caracterização genômica de vírus recombinantes HIV-1 BC

(artigo: *Genomic characterization of HIV-1 BC recombinant Viruses*).

O HIV-1 é caracterizado por um alto índice de diversidade genética, representada por grupos, subtipos e formas recombinantes. Até o momento, foram identificadas 14 Formas Recombinantes Circulantes (CRFs) com mosaicos dos subtipos B e C, três delas em 2021: CRF108\_BC na Espanha, CRF110\_BC e CRF118\_BC na China. Análises de recombinação e filogenética foram conduzidas em 225 amostras BC encontradas em banco de dados usando ferramentas de bioinformática. As amostras foram obtidas de 17 países entre 1992 e 2018. Quatorze CRFs e 82 padrões distintos de recombinação de Formas Recombinantes Únicas (URFs) foram identificados. Pontos preferenciais de recombinação foram encontrados nos genes pol (66,7%) e env (62,5%), enquanto o subtipo C é a forma genética predominante nos genes gag, vif, vpr, tat, rev, tat2, rev2, nef e 3'-LTR regiões. Posições específicas nos genes pol e env mantêm o subtipo B, o que pode estar relacionado à aquisição de vantagem adaptativa sobre o subtipo C.

*O referido trabalho será submetido para publicação.*

## GENOMIC CHARACTERIZATION OF HIV-1 BC RECOMBINANT VIRUSES

Rodrigo Cunha Oliveira<sup>1</sup> and Joana Paixão Monteiro-Cunha<sup>1\*</sup>

<sup>1</sup> Universidade Federal da Bahia, Departamento de Bioquímica e Biofísica, Salvador, Bahia, Brasil

\*Correspondence: Joana Paixão Monteiro Cunha; [joana.monteiro@ufba.br](mailto:joana.monteiro@ufba.br)

**KEYWORDS:** HIV-1; Recombination; BC; CRF.

### ABSTRACT

The Human Immunodeficiency Virus 1 (HIV-1) is characterized by a high rate of genetic diversity, represented by groups, subtypes and recombinant forms. To date, 14 Circulating Recombinant Forms (CRFs) with subtypes B and C mosaics have been identified, three of them in 2021: CRF108\_BC in Spain, CRF110\_BC and CRF118\_BC in China. This research investigated the genetic diversity among circulating BC recombinant viruses. Near-Full Length Genomes (NFLG) of 225 BC strains were retrieved from the Los Alamos database. Recombination and phylogenetics analysis were conducted using bioinformatics tools. Samples were obtained from 17 countries between 1992 and 2018. Fourteen CRFs and 82 distinct Unique Recombinant Forms (URFs) recombination patterns were identified. Preferential points of recombination were found in *pol* (66.7%) and *env* (62.5%) genes, while subtype C is the predominant genetic form in *gag*, *vif*, *vpr*, *tat*, *rev*, *tat2*, *rev2*, *nef* and 3'-LTR genomic regions. Specific positions in *pol* and *env* genes maintain the B subtype, which may be related to the acquirin of adaptive advantage over subtype C.

## 1. INTRODUCTION

The Human Immunodeficiency Virus (HIV) discovered in the early 1980s is classified into two types of viruses: HIV-1 and HIV-2 (Ioachim, 1990; Robertson et al., 2011). HIV-1 is highly diversified, being represented by four groups (M, N, O and P). The M group is further classified into 10 subtypes (A, B, C, D, F, G, H, J, K and L), sub-subtypes (A1-A8 and F1, F2) and Circulating or Unique Recombinant Forms (CRF and URF) (Robertson et al., 2011; Yamaguchi et al., 2019).

The subtype C accounts for nearly one-half of HIV-1 infections worldwide (46.6%), whereas the B subtype is globally distributed (Hemelaar et al., 2019). Currently, 14 CRFs (07, 08, 31, 57, 60, 61, 62, 64, 85, 86, 88, 108, 110 and 118) involving both subtypes are described in the Los Alamos Laboratory (LANL) database. The first (CRF07\_BC) was described in 2000 (Su et al., 2000) and the last (CRF118\_BC) in 2021 (Gao et al., 2021). Most of BC CRFs and URFs have been reported in China (<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). However, some CRFs were also identified in Brazil and Italy (Santos et al., 2006; Simonetti et al., 2014). Molecular epidemiology analyses show that the frequency of recombinant forms of HIV-1 in the world was 22.9% between 2010 and 2015 (Hemelaar et al., 2019). Furthermore, when evaluating countries individually, an increase in the number of virus variants with high levels of CRF and URF was reported (Hemelaar et al., 2020). In Brazil, the subtype B is the prevalent genetic form, followed by the F1 and C subtypes. Therefore, recombinant genotypes (BF1 and BC) are frequently reported throughout this country (Arruda et al., 2018). The CRF31\_BC is found mainly in the southern region of Brazil (8.1%), whereas in the Southeast and Central-West this form accounts for less than 1% (Arruda et al., 2018). Meanwhile, in the northern (1.1%) and northeastern (4%) regions of Brazil, only BC URFs have been detected (Chaves et al., 2021; Delatorre et al., 2017).

Although the UNAIDS report (UNAIDS, 2022) showed lower rates of new HIV infections in the previous year, new genetic forms are still emerging, for instance, CRF108\_BC in Spain (Cañada et al., 2021), CRF110\_BC and CRF118\_BC in China (Gao et al., 2021; Li et al., 2021). Due to the expansion of new HIV-1 genotypes, this research investigated the genetic diversity among BC recombinant strains circulating worldwide using viral nucleotide sequences, mapping preferential points of recombination through recombination detection programs and discussing possible advantages of the maintenance of pure subtypes in the different genomic regions.

## 2. MATERIALS AND METHODS

Two hundred and twenty-five Near-Full Length Genomes (NFLG) sequences (>7000 base pairs [bp]) identified as BC recombinants were collected from LANL database (<http://www.hiv.lanl.gov>) until December 2021. The sequences were cataloged based on GenBank accession number, CRF classification, country and year of sample collection (when available). In order to investigate and compare the recombination patterns, sequences were analyzed using jpHMM online tool (Schultz et al., 2009), SimPlot (Bootscan method) (Salminen et al., 1996), using the following parameters: 400 bp of slider window size and 40 bp steps size and RDP5 software (RDP method) (Martin et al., 2021). The frequency of pure subtype B, pure subtype C and recombination (R) within the HIV-1 genes was investigated among distinct recombinant patterns (Souza et al., 2019). The intervals for recombination breakpoints based on HXB2 numbering obtained for each sequence were used to identify the regions where the recombination between B and C subtypes were more frequent. In addition, we compared this breakpoint against the Simplot and RDP5 plot range. The frequency of recombination was evaluated at each interval of 100 nucleotides and calculated as the number of distinct recombinant patterns that showed a recombination point in that region among the total of patterns. A single viral sequence for each recombination pattern was considered in this analysis.

In order to investigate the phylogenetic relationships, BC recombinant sequences were aligned with 42 reference sequences of HIV-1 pure subtypes obtained from LANL database. The alignment was performed using MAFFT online program (Kato et al., 2005) (<http://mafft.cbrc.jp>) under the command: `mafft --thread 8 --threadtb 5 --threadit 0 --reorder --auto input > output` and manually edited using BioEdit software (Hall, 1999). The final dataset, composed by 267 sequences (7008bp), was assessed for the presence of phylogenetic signal by applying the likelihood mapping analysis implemented in the TreePuzzle Program (Schmidt, et al., 2003). The percentage of fully resolved probability mapping quartets totals 99.6%. The Maximum likelihood (ML) tree was reconstructed using IQ-TREE 1.6.8 webserver (Command: `path_to_iqtree -s aln.fasta -m TEST -bb 1000 -alrt 1000`) (Trifinopoulos et al., 2016). The analysis was performed under the GTR+I+G+F nucleotide substitution model, which was inferred in Modelfinder (Kalyaanamoorthy et al., 2017). The reliability of each cluster was evaluated by analyzing 1,000 Bootstrap (BS) replicates and with the approximate likelihood-ratio test (aLRT) based on the Shimodaira-Hasegawa-like procedure (Anisimova et al., 2006).

The ML tree was visualized with the FigTree program version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 3. RESULTS

The 225 NFLG sequences of HIV-1 BC recombinant strains were collected between 1992 and 2018 (supplementary material 1). These variants have been isolated in 17 countries and four continents (Africa, America, Asia and Europe) (Figure 1A). One hundred thirty-five (60.0%) sequences (Figure 1B) were classified among 14 distinct CRFs (Figure 2). Together, CRF07\_BC (n=43) and CRF08\_BC (n= 35) accounts for almost 60% of CRFs genomes, being reported mostly in China whereas the Korean Republic, Taiwan, United Kingdom and Vietnam had the lowest rates. Furthermore, CRF57\_BC (n=7), CRF61\_BC (n= 4), CRF62\_BC (n= 3), CRF64\_BC (n= 9), CRF85\_BC (n= 11), CRF86\_BC (n= 3), CRF88\_BC (n= 3), CRF110\_BC (n= 3) and CRF118\_BC (n= 1) were only found in China. Likewise, the CRF31\_BC (n= 3) was exclusively reported in Brazil and CFR108\_BC (n= 5) in Spain, while CRF60\_BC (n= 5) has been identified in Brazil, Italy and France. Ninety (40%) NFLG sequences were classified as URFs (Figure 1B). Among them, 82 different recombination patterns were observed (<https://data.mendeley.com/datasets/gxbx87bg77/2>). Three sequences presented distinct subtype assignment, being classified as pure subtype C, complex recombinant and partially unassigned. Five sequences were clones or presented the same recombination pattern.

Overall, 96 recombination patterns between subtypes B and C were found circulating worldwide (14 CRFs + 82 URFs). Considering a single sequence for each recombination pattern, the frequency of subtype B, C and intragenic recombination was assessed in each genomic region (Figure 3A). Recombination occurred more frequently in *pol* (66.7%) and *env* (62.5%) structural genes, while the only one regulatory gene with high rate of recombination (>30%) was *nef*. Through recombination analyses, the nucleotide intervals where recombination occurs more frequently were investigated in these three genes. Based on the HXB2 reference strain, in the *pol* gene, the events were more frequent at positions 3100-3200bp (25%; Reverse Transcriptase) and 4,900-5,000bp (20%; Integrase/VIF protein). In *env* and *nef* genes, positions 6,300-6,400bp (22%; gp120), 8,500-8,600bp (22%), 8,700-8,800bp (24%) (gp41/NEF protein) and 8,900-9,000bp (40% NEF protein) were found as preferred recombination points (Figure 3B). In all other HIV-1 genes, pure subtype C (25% - 79.2%) predominated over the pure subtype B (4.2% - 33.3%) (Figure 3A).

The phylogenetic relationships between the viral strains are represented in the ML tree (Figure 4). All BC recombinants (n=225) grouped together with reference sequences of subtypes B and C (aLRT=99/BS=100). Sequences classified as CRF07\_BC, CRF85\_BC and 16 Chinese URFs plus one URF from Vietnam branched together in the same clade (aLRT=98/BS=91). Recombinant samples from Dominican Republic, India, South Africa and Sweden grouped with subtype C reference sequences from Ethiopia, Indian and South Africa (aLRT=97/BS=97), in the same node with CRF60\_BC, CRF31\_BC and CRF108\_BC clustered with 13 URF from South America. At last, sequences of CRF57\_BC, CRF\_61, CRF62\_BC, CRF64\_BC, CRF86\_BC, CRF88\_BC, CRF110\_BC and CRF118\_BC were presented intermixed with URFs from Asia (China, India, Myanmar and Vietnam) and Europe (United Kingdom). The base of the clade containing the BC sequences is composed of URFs with a predominance of subtype B (Brazil and Cyprus) along the genome with smaller C insertions plus subtype B reference sequences (aLRT: 100 /BS: 100).

#### 4. DISCUSSION

The extensive genetic diversity of HIV-1 is due to its high mutation and replication rates associated with the existence of two copies of single-stranded RNA. During reverse transcription of viral replication, both RNA copies are used for strand-transfer mediated recombination, allowing for rapid evolution under immunological and therapeutic pressures. In fact, this remarkable variability of HIV-1 strains is one of the greatest challenges for vaccine development (Perelson et al., 1996). Thus, studies on the genetic diversity of HIV-1 are important for understanding the molecular mechanisms of viral evolution and their significance for disease and infectivity conditions (Mansky, 1998).

Nearly 76% of 225 BC recombinant strains were isolated in China. Of the 14 BC CRFs recognized so far, 11 circulate in China (Figure 1A), in addition to 50 BC URFs with distinct recombination patterns. This indicates that recombination events are occurring in a high frequency in that country. The HIV-1 epidemic in China reached a high rate in the late 80's and the lineage B' (Thailand) and the subtype C from India were the dominant genotypes (Heng et al., 1994). In China, the high numbers of HIV cases were associated with Injecting Drug Users (IDUs) and sexual contact (Li et al., 2020; Piyasirisilp et al., 2000). These factors may have been the major cause of the emergence of the new genotypes that included the first BC CRFs: CRF\_07 and CRF\_08 (Feng et al., 2016). The wide variety of BC recombinant patterns in China

is most likely related to second-generation recombination events. In these cases, the same breakpoint is shared between the recombinant (CRF07 and CRF08) or pure subtype (B or C) parental sequences, leading to the rise of the URFs over time (Han et al., 2013). Some evidence of this mechanism was identified in four new BC URFs presenting recombinant structures equivalent to CRF\_07 and CRF\_85 (Wang et al., 2017; Zhang et al., 2018; Zhang et al., 2019). Furthermore, two new BC CRFs were described in 2021: CRF\_110 (Li et al., 2021) and CRF\_118 (Gao et al., 2021), both found in IDUs. These HIV-1 variants present subtype B insertions in *pol*, *env* and *nef* (CRF\_110) and *gag*, *pol*, *vpu*, *env*, *rev2* and *nef* genes (CRF\_118) (Figure 2) and were phylogenetically related to Chinese URFs in the ML analysis (Supplementary material 2). Piyasirisilp and collaborators, (2000), investigated BC samples from different locations in China and reported a low diversity rate among these sequences, which could be related to the existence of selective pressure for the maintenance of subtype B' insertions on the C genome, mainly at the level of Reverse Transcriptase (RT). Based on bioinformatics assays, the authors found a high degree of conservation of Aspartate and Lysine in the RT active site in the subtype B' genotype, which could be related to the enhancement of the enzymatic activity, influencing the reverse transcription error rate, leading to a slower accumulation of mutations per replication cycle. Thus, the fragment would be transmitted between the strains through positive selection, since the studied samples had distinct phylogenetic origins. On the other hand, a molecular characterization in the gp120 envelope glycoprotein, among old and new CRF\_07 isolates, led to the conclusion that the new strains show more potential N-glycosylation sites than do the oldest ones.<sup>36</sup> These shared genetic characteristics provide evidence on the evolution of the glycan shield to escape from the antibody neutralization - an important factor that may be involved with the rapid spread of CRF\_07 (Meng et al., 2007).

A low frequency of the B subtype throughout the HIV-1 genome (>30% only in *vpu*) was observed among BC recombinant sequences (Figure 3A), while subtype C (>30% in *gag*, *vif*, *vpr*, *tat*, *rev*, *tat2*, *rev2*, *nef* and 3'-LTR) and recombination (> 50% in *pol*, *env* and *nef*) were more frequent. Concerning the structural genes, recombination events occurred more frequently in *pol* (66.7%) and *env* (62.5%), whereas *gag* presented only 36.5% of intrasubtype recombination. This pattern was not observed in most accessory and regulatory genes, which maintained subtype C in most samples, with a recombination rate of less than 20%. Conversely, the *nef* gene presented a high rate of recombination (38.5%). Therefore, among 96 distinct BC recombination patterns, preferential points of recombination were found in *pol*, *env* and *nef*

genes (Figure 3B). In comparison, among over 100 BF different patterns, recombination occurred more frequently in all structural genes (*gag* [69.3%], *pol* [84.9%] and *env* [51.4 %]), while *nef* showed a lower rate of intragenic recombination (25.5%) (Souza et al., 2019). Furthermore, among BC recombinant patterns, the subtype C was more frequent throughout the viral genome, especially in the accessory and regulatory genes, while among BF mosaics, the subtype B was predominant in those genes, except in *vif* (Souza et al., 2019).

The NEF protein is involved in the replicative ability of HIV-1 (Foster et al., 2008) and is also highly conserved in CRF07 (Meng et al., 2007). Here, 43/82 URF patterns and 8/14 CRF patterns (07, 61, 64, 85, 86, 108, 110 and 118) show a B fragment in the *nef* gene while CRF08 and CRF60 are fully B. Overall, the NEF protein has the function of disabling cellular antiviral mechanisms, through the downregulation of CD4<sup>+</sup> and MHC-1 in immune system cells (Mann et al., 2015). Nevertheless, there are differences in the response generated by viral isolates of different subtypes. For instance, subtype B has shown a greater ability to downregulate MHC-1 molecules compared to subtype C, possibly enhancing cytotoxic activity evasion (Mann et al., 2015). The transmembrane proteins Serine Incorporator 3 and 5 (SERINC-3 and SERINC-5) are also targets of the NEF protein activity (Usami et al., 2015). Both molecules have antiviral activity, with the function of preventing the binding between the virus and the target cell. Between the two types, SERINC-5 is more effective against HIV. When the NEF protein interacts by myristoylation in the inner cell side, it internalizes the SERINCs (Usami et al., 2015). A molecular comparative prediction analysis between subtypes revealed that subtype C has a lower capacity to internalize SERINC-5, unlike subtype B which is efficient in neutralizing this antiviral factor (Jin et al., 2020).

After China, Brazil shows the highest number of BC recombinant viruses, with two identified CRFs (CRF31 and CRF60) and 14 URFs. Most of these BC samples were isolated in southern Brazil, where subtypes B and C co-circulate. Passaes and collaborators, (2009), proposed that second-generation recombination events were also involved in BC strain diversity and the structural genes *gag*, *pol* and *env* were considered as hotspots for recombination events (Passaes et al., 2009).

Interestingly, most of the BF recombinant strains circulating worldwide have been isolated in Brazil (51.3%), where nine out of 15 BF CRFs have been described.<sup>19</sup> Taken together, these data show the high occurrence of recombination events in Brazil and China, as discussed above. In fact, high recombination rates have been reported in places where more than one subtype co-circulates. This is favored through the multiple subtypes co-infections and transmission



pathway, that selects the strains. (Hemelaar et al., 2020; Kiwelu et al., 2013).

Five sequences are classified as CRF60 (Los Alamos), three of them from Italy (Simonetti et al., 2014) and one from Brazil (Sanabani et al., 2011) and France (Frange et al., 2008). The sample from Brazil was not identified in the report in 2011. The authors found two BC samples, and one of them was CRF31 (Sanabani et al., 2011).

Regarding the French CRF60, the authors described it as a URF between the genotype B and Brazilian subtype C. The sequence mosaic revealed B insertions in the Reverse Transcriptase and in the Integrase, similar to the CRF31 and patterns of Brazilian URFs (Frange et al., 2008). Moreover, phylogenetic analyzes reveal a close relationship between Italian CRF60 with pure subtypes B and C from South American. The hypotheses are based that CRF60 emerged in South America continent or in Italy through the subtype C introduced by travelers (Simonetti et al., 2014).

Another evidence of the probable contribution of South America to the European epidemic is the CRF108 Spanish. The subtype C Brazilian sequences are presented as ancestors of the CRF108 cluster (Cañada et al., 2021). In our phylogenetic tree, we provide more evidence of this relationship, being all the CRFs 31, 60 and 108 groups with 12 URFs from South America (Argentina, Brazil and Uruguay) plus two reference sequences from Brazil and Ethiopia (BS: 99/ aLRT: 96). Among the 11 samples from India, one of them shared a structure similar to the sequence from Argentina (Lakhase et al., 2008). Phylogenetic analyses revealed that although the B insertions in the env gene are similar, the evolutionary origin of both recombinants is different, suggesting that URF Argentina did indeed arise in South America (Aulicino et al., 2018; Lakhase et al., 2008).

Subtype recombination is one of the main mechanisms that result in rapid evolution and genetic diversity (Zyl et al., 2018). According to Li and colleagues, 2020, the CRF\_07 still spreads, mainly in China. Besides, the high rates of BC URFs in that country associated with CRF\_07 transmission can lead to the emergence of new genotypes (CRF or URF) through the second-generation recombination effect. Furthermore, Takebe and collaborators, (2003), suggest that there are mechanisms allowing the convergent evolution of the recombinants with similar breakpoints, even when they have different origins. Thereby, these factors may grant the virus selective advantages, driving recombination events, which may provide an opportunity to produce more fitness compared to parental viruses (Yang et al., 2003).

## 5. CONCLUSION

To conclude, a great diversity of recombination patterns between subtypes B and C was detected, with 96 distinct profiles among 225 recombinant samples. These isolates are found mainly in Brazil and China, with a high number of CRFs (China) and URFs (Brazil and China), most likely due the B and C subtypes co-circulation. In this study, we highlighted the genes with higher recombination rates (*pol and env*) and the preferential maintenance of the subtype C among the sequences. Based on the frequency of B insertions, we discussed some characteristics that may be transmitted among BC strains, which could favor their spread.

### **Acknowledgements**

R.C.O. is greatfull to Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for the Phd. fellowship.

### **Conflict of interest**

There is no conflict of interest among co-authors.

### **Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### **Authors contributions**

R.C.O. Conceptualisation, Investigation, Writing - Original Draft Preparation;  
J.P.MC. Conceptualisation, Validation, Investigation, Writing – Original Draft Preparation,  
Supervision

### **References**

- Anisimova, M., Gascuela, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* 55 (4): 539–552. <https://doi.org/doi:10.1080/1063515060075545355>
- Arruda, M.B., Boulosa, L.T., Cardoso, et al. 2018. Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): a survey of treatment-naive individuals. *Journal of the International AIDS Society.* 21 (3): e25032. <https://doi.org/doi/10.1002/jia2.25032/full>
- Aulicino, P.C., Kopka, J., Rocco, C., et al. 2005. Sequence Analysis of a South American HIV

- Type 1 BC Recombinant. *AIDS Research and Human Retroviruses*. 21(10): 894–896. <https://doi.org/doi:10.1089/aid.2005.21.894>
- Cañada, J.E., Delgado, E., Gil, H., et al. 2021. Identification of a New HIV-1 BC Intersubtype Circulating Recombinant Form (CRF108\_BC) in Spain. *Viruses*. 13 (1): 93. <https://doi.org/10.3390/v13010093>.
- Chaves, Y.O., Pereira, F.R., Pinheiro, R.S., et al. 2021. High Detection Rate of HIV Drug Resistance Mutations among Patients Who Fail Combined Antiretroviral Therapy in Manaus, Brazil. *Biomedical Research International*. 5567332. <https://doi.org/10.1155/2021/5567332>
- Delatorre, E., Couto-Fernandez, J.C., Bello, G., 2017. HIV-1 Genetic Diversity in Northeastern Brazil: High Prevalence of Non-B Subtypes. *AIDS Research and Human Retroviruses*. 33 (7): 639-647. <https://doi.org/10.1089/AID.2017.0045>.
- Feng, Y., Takebe, Y., Wei, H., et al. 2016. Geographic origin and evolutionary history of China's two predominant HIV-1 circulating recombinant forms, CRF07\_BC and CRF08\_BC. *Scientific Reports*. 6: 19279. <https://doi.org/10.1038/srep19279>
- Foster, J.L., Garcia, J.V., 2008. HIV-1 Nef: at the crossroads. *Retrovirology*. 5: 84. <https://doi.org/10.1186/1742-4690-1185-1184>.
- Frange, P., Galimand, J., Vidal, N., et al. 2008. New and old complex recombinant HIV-1 strains among patients with primary infection in 1996–2006 in France: The French ANRS CO06 primo cohort study. *Retrovirology*. 5: 69. <https://doi.org/10.1186/1742-4690-5-69>
- Gao, Q., Feng, Y., Gao, L., et al. 2021. Genetic characteristics of a New HIV-1 subtype B/C intersubtype circulating recombinant form (CRF118\_BC) identified in Yunnan, China. *Journal of Infection*. 83 (2): 237-279. <https://doi.org/10.1016/j.jinf.2021.05.006>
- Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41: 95-98. <https://doi.org/citeulike-article-id:691774>
- Han, X., An, M., Zhao, B., et al. 2013. High Prevalence of HIV-1 Intersubtype B9/C Recombinants among Injecting Drug Users in Dehong, China. *PLoS One*. 31: 8 (5) e65337. <https://doi.org/10.1371/journal.pone.0065337>
- Hemelaar, J., Loganathan, S., Elangovan, R., et al. 2019. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infectious Disease*. 19 (2): 143-155. [https://doi.org/10.1016/S1473-3099\(18\)30647-9](https://doi.org/10.1016/S1473-3099(18)30647-9)
- Hemelaar, J., Loganathan, S., Elangovan, R., et al. 2020. Country Level Diversity of the HIV-1 Pandemic between 1990 and 2015. *Journal of Virology*. 95 (2): e01580-20. <https://doi.org/10.1128/JVI.01580-20>
- Heng, X., Tian, C., Choi, K.H., et al. 1994. Injecting drug use and HIV infection in southwest China. *AIDS*. 8: 1141–1147.

- Ioachim, H.L. 1990. Immunopathogenesis of Human Immunodeficiency Virus infection. *Cancer Research*. 50: 5612-5617.
- Jin, S.W., Mwimanzi, F.M., Mann, J.K., et al. 2020. Variation in HIV-1 Nef function within and among viral subtypes reveals genetically separable antagonism of SERINC3 and SERINC5. *PLoS Pathogens*. 16(9): e1008813. <https://doi.org/10.1371/journal.ppat.1008813>
- Kalyaanamoorthy, S., Minh, B., Wong, T., et al. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*. 14 (6): 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., Kuma, K., Toh, H., et al. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*. 33 (2): 511–518. <https://doi.org/10.1093/nar/gki198>
- Kiwelu, I. E., Novitsky, V., Margolin, L., et al. 2013. Frequent Intra-Subtype Recombination among HIV-1 Circulating in Tanzania. *Plos One*. 8 (8), e71131. <https://doi.org/10.1371/journal.pone.0071131>
- Lakhase, S., Tripathy, S., Paranjape, R., et al. 2008. Characterization of B/C recombinants of near full-length HIV type 1 from northeastern India with mosaics identical to ARE195FL but with a different ancestral origin. *AIDS Res Hum Retroviruses*. 24 (1): 92-99. <https://doi.org/doi:10.1089/aid.2007.0214>.
- Li, W., Li, X., He, Y., et al. 2020. The evolutionary and transmission characteristic of HIV-1 CRF07\_BC in Nanjing, Jiangsu. *Journal Medical of Virology*. 10. <https://doi.org/doi:10.1002/jmv.25854>
- Li, J., Gao, Q., Zhang, M., et al. 2021. A newly emerging HIV-1 circulating recombinant form (CRF110\_BC) comprising subtype B and C among intravenous drug users in Yunnan, China. *Journal of Infection*. 82 (3): e8–e10. <https://doi.org/10.1016/j.jinf.2020.12.015>
- Lole, K.S., Bollinger, R.C., Paranjape, R.S., et al. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*. Jan; 73 (1):152-160. <https://doi.org/10.1128/jvi.73.1.152-160.1999>
- Mann, J.K., Omarjee, S., Khumalo, P., et al. 2015. Genetic determinants of Nef-mediated CD4 and HLA class I down-regulation differences between HIV-1 subtypes B and C. *Virology Journal*. 12: 200. <https://doi.org/10.1186/s12985-015-0429-7>
- Mansky, L.M. 1998. Retrovirus mutation rates and their role in genetic variation. *Journal of General Virology*. 79, 1337–1345.
- Martin, D. P., Varsani, A., Roumagnac, et al. 2021 RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*. 7 (1). <https://doi.org/10.1093/ve/veaa087>

- Meng, Z., Xing, H., Ma, L., et al. 2007. Genetic characterization of three newly isolated CRF07\_BC near full-length genomes in China. *AIDS Research and Human Retroviruses*. 23 (8): 1049-1054. <https://doi.org/10.1089/aid.2007.0058>
- Nadai, Y., Eyzaguirre, L.M., Sill, A., et al. 2009. HIV-1 epidemic in the Caribbean is dominated by subtype B. *PLoS One*. 4 (3): e4814. <https://doi.org/10.1371/journal.pone.0004814>
- Passaes, C.P.B., Guimarães, M.L., Bello, G., et al. 2009. Near Full-Length Genome Characterization of HIV Type 1 Unique BC Recombinant Forms from Southern Brazil. *Aids Research and Human Retroviruses*. 25: 12. <https://doi.org/10.1089=aid.2009.0167>
- Perelson, A.S., Neumann, A.U., Markowitz, M., et al. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 271(5255):1582–6.
- Piyasirisilp, S., McCutchan, F.E., Carr, J.K., et al. 2000. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *Journal of Virology*. 74 (23): 11286-11295. <https://doi.org/10.1128/jvi.74.23.11286-11295.2000>.
- Robertson, D.L., Anderson, J.P., Bradac, J.A., 2011. HIV-1 nomenclature proposal. *Science*. 288 (5463): 55-56. <https://doi.org/10.1126/science.288.5463.55d>
- Sanabani, S.S., Pastena, E.R.S., Costa, A.C., et al. 2011. Characterization of Partial and Near Full-Length Genomes of HIV-1 Strains Sampled from Recently Infected Individuals in São Paulo, Brazil. *Plos One*. 6: 10. e25869. <https://doi.org/10.1371/journal.pone.0025869>
- Santos, A.F., Sousa, T.M., Soares, E.A.J.M., et al. 2006. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS*. 20; (16): 2011-2019. <https://doi.org/10.1097/01.aids.0000247573.95880.db>.
- Schmidt, H.A., Petzold, E., Vingron, M., et al. 2003 Molecular Phylogenetics: Parallelized Parameter Estimation and Quartet Puzzling. *J. Parallel Distrib. Computacional*; 63: 719–727. [https://doi.org/10.1016/S0743-7315\(03\)00129-1](https://doi.org/10.1016/S0743-7315(03)00129-1)
- Schultz, A.K., Zang, M., Bulla, I., et al. 2009. jpHMM: Improving the reability of recombination prediction in HIV-1. *Nucleic Acids Research*. 37 (Web Server): W647–W651. <https://doi.org/10.1093/nar/gkp371>
- Souza, J.S.M., Júnior, J.J.S., Brites, C., et al. (2019) Molecular and geographic characterization of hiv-1 bf recombinant viruses. *Virus Research*. 270: 197650. <https://doi.org/j.virusres.2019.197650>
- Simonetti, F.R., Lai, A., Monno, L., et al. 2014. Identification of a new HIV-1 BC circulating recombinant form (CRF60\_BC) in Italian young men having sex with men. *Infection, Genetics and Evolution*. 23: 176–181. <https://doi.org/10.1016/j.meegid.2014.02.007>
- Su, L., Graf, M., Zhang, Y., et al. 2000. Characterization of a Virtually Full-Length Human Immunodeficiency Virus Type 1 Genome of a Prevalent Intersubtype (C/B') Recombinant

Strain in China. *Journal of Virology*. 74 (23): 11367–11376. <https://doi.org/10.1128/jvi.74.23.11367-11376.2000>

Takebe, Y., Motomura, K., Tatsumi, M., et al. 2003. High prevalence of diverse forms of HIV-1 intersubtype recombinants in Central Myanmar: geographical hot spot of extensive recombination. *AIDS*. 26; 17 (14) :2077-2087. <https://doi.org/10.1097/00002030-200309260-00009>.

Trifinopoulos, J., Nguyen, L.T., Haeseler, A.V., et al. 2016. W-IQ-TREE: a fast-online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*. 44 (W1): W232–W235. <https://doi.org/10.1093/nar/gkw256>

UNAIDS. UNAIDS Data 2022. ([https://www.unaids.org/sites/default/files/media\\_asset/data-book-2022](https://www.unaids.org/sites/default/files/media_asset/data-book-2022). Accessed in May 24th, 2023).

Usami, Y., Wu, Y., Gottlinger, H.G., 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*. 526 (7572): 218–223. <https://doi.org/10.1038/nature15400>

Wang, X., Zhang, M., Li, J., et al. 2017. Genetic Characterization of a Unique Recombinant strain identified in Yunnan with genome comprised of B and C. *AIDS Research and Human Retroviruses*, 33 (6): 614-620. <https://doi.org/10.1089/AID.2017.0013>.

Yamaguchi, J., McArthur, C., Vallari, A., et al. 2019. Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *Journal of Acquired Immune Deficiency Syndromes*. 83: (3); 319-322. <https://doi.org/10.1097/qai.0000000000002246>

Yang, R., Kusagawa, S., Zhang, C., et al. 2003. Identification and characterization of a new class of human immunodeficiency virus type 1 recombinants comprised of two circulating recombinant forms, CRF07\_BC and CRF08\_BC, in China. *Journal of Virology*. 77 (1); 685-695. <https://doi.org/10.1128/jvi.77.1.685-695.2003>.

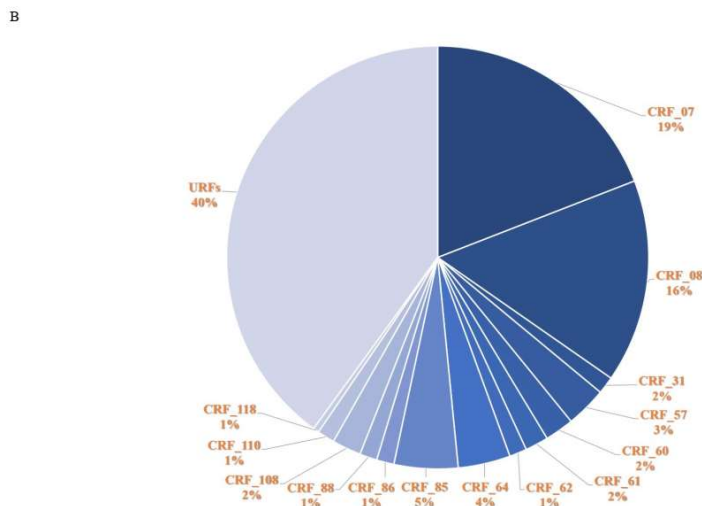
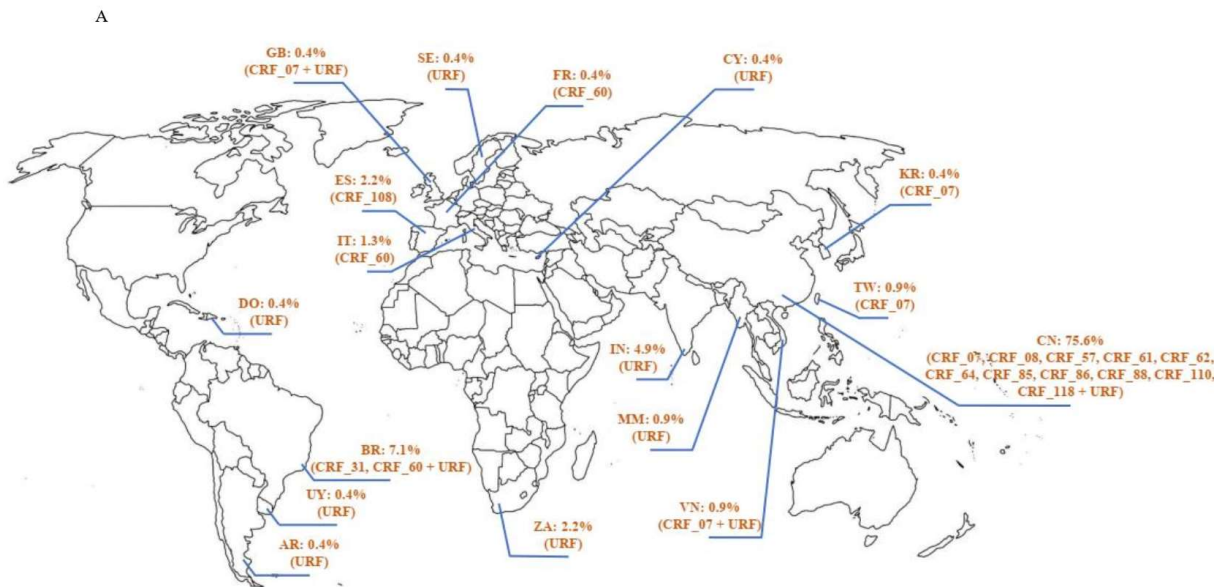
Zhang, M., Xin, R., Chang, W., 2018. Identification of a Novel HIV-1 Recombinant Form (B/C) in Men Who Have Sex with Men in Shaanxi, China. *AIDS Research and Human Retroviruses*. 34 (5): 467-470. <https://doi.org/10.1089/AID.2018.0003>

Zhang, M., Li, H., Zou, Y., et al. 2019. Identification of Three HIV-1 Unique (B/C) Recombinant Forms in Shaanxi, China. *AIDS Research and Human Retroviruses*. 35 (9): 876-880. <https://doi.org/10.1089/AID.2019.0108>

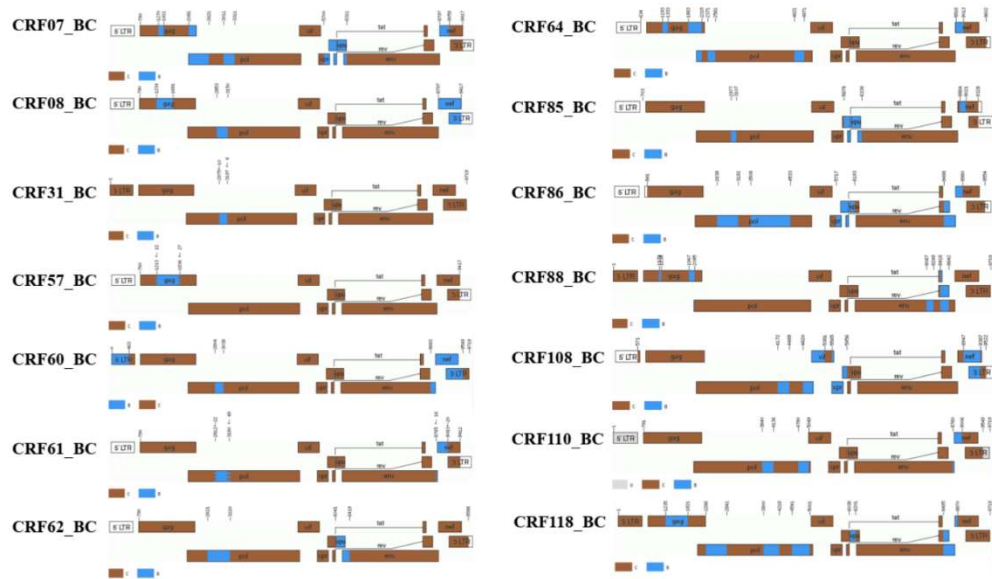
Zyl, G.V., Bale, M.J., Kearney, M.F. 2018. HIV evolution and diversity in ART-treated patients. *Retrovirology*. 15: 14. Published online. <https://doi.org/10.1186/s12977-018-0395-4>

**FIGURE CAPTIONS**

**Figure 1: A** - Global distribution HIV-1 BC recombinants strains. Genetic forms found in each country are indicated in parentheses with the respectively frequency data. Countries are indicated by a two-letter code: AR: Argentina, BR: Brazil, CN: China, CY: Cyprus, DO: Dominican Republic, ES: Spain, FR: France, GB: United Kingdom, IN: India, IT: Italy, KR: Korean Republic, MM: Myanmar, SE: Sweden, TW: Taiwan, UY: Uruguay, VN: Vietnam and ZA: South Africa. **B** - Frequencies among BC circulating and unique recombinant viruses. URF: Unique Recombinant Forms. CRF: Circulating Recombinant Form.

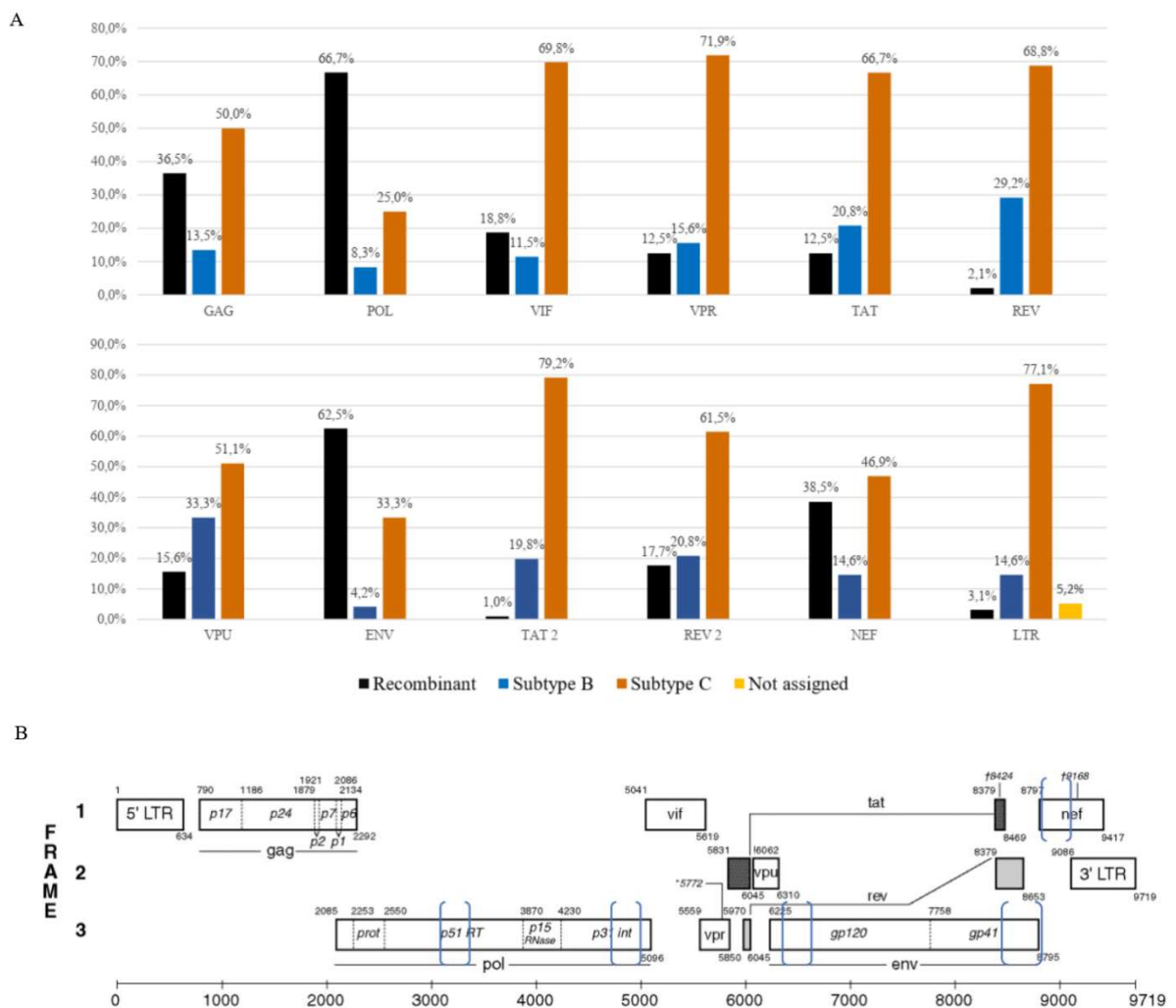


**Figure 2:** Recombination patterns of the 14 CRF\_BC circulating worldwide. Adapted from Los Alamos HIV sequence database (<https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

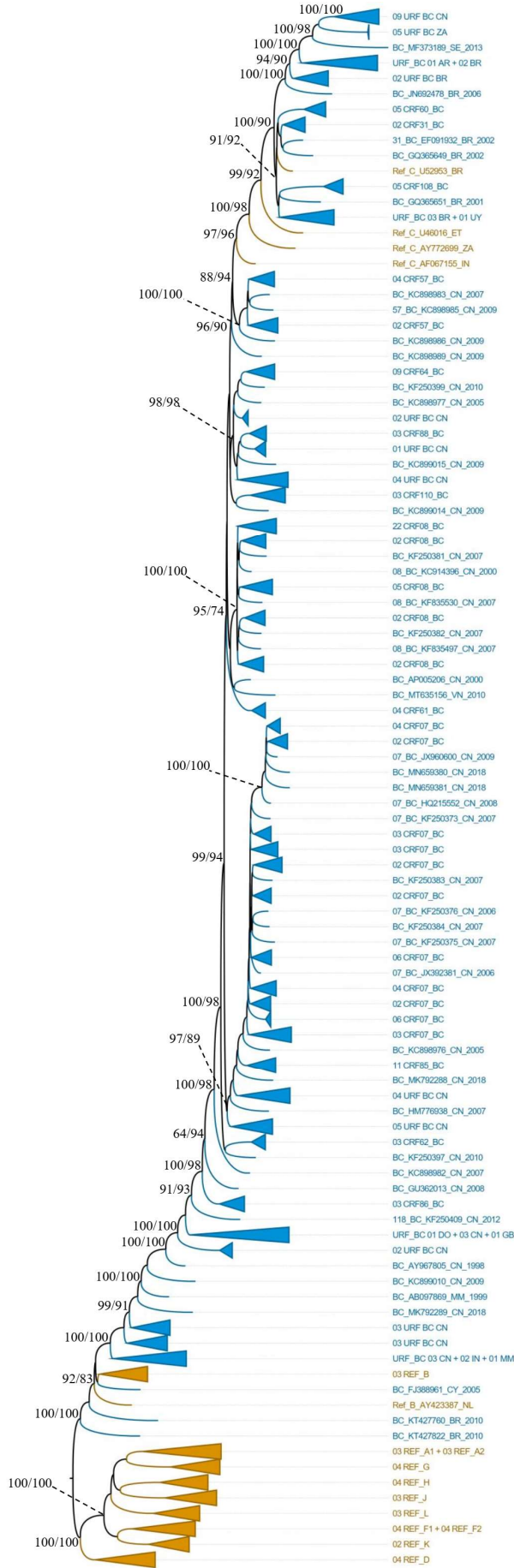




**Figure 3:** Preferential points of recombination in the HIV-1 genome of BC variants. **A** - Pure subtype and recombination frequencies within each HIV gene among BC recombinant viruses (LTR not assigned: without biological information in available nucleotide sequences). **B** - Specific nucleotide positions in *pol*, *env* and *nef* genes with recombination points  $\geq 20\%$  of the analyzed sequences are represented between blue lines.



**Figure 4:** Maximum likelihood analysis of HIV-1 subtype BC sequences (7008bp) rooted by the midpoint. The blue branches represent the BC samples and the orange represent the subtype reference sequences. In the collapsed clusters, the subtype classification (CRF or URF), country and number of sequences. The countries are indicated by two letters code: AR: Argentina, BR: Brazil, CN: China, CY: Cyprus, DO: Dominican Republic, ES: Spain, FR: France, GB: United Kingdom, IN: India, IT: Italy, KR: Korean Republic, MM: Myanmar, SE: Sweden, TW: Taiwan, UY: Uruguay, VN: Vietnam and ZA: South Africa. The statistical support is indicated as aLRT/bootstrap values. The tree was built under the evolutionary model GTR + I + G + F and visualized in the Figtree v1.4.4 software. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating 0.07 nucleotide substitutions per site.



**Supplementary material 1: Nucleotide sequence list**

ACCESS NUMBER	SUBTYPE CLASSIFICATION	COUNTRY	YEAR
AB097869	BC	MM	1999
AB097873	BC	MM	1999
AB746342	08_BC	CN	2001
AB746343	08_BC	CN	2001
AB746344	08_BC	CN	2000
AB746345	08_BC	CN	2000
AB773884	08_BC	CN	2000
AB773885	08_BC	CN	2000
AF286226	07_BC	CN	1997
AF286229	08_BC	CN	1998
AF286230	07_BC	CN	1998
AF503396	07_BC	CN	-
AP005206	BC	CN	2000
AY008715	08_BC	CN	1997
AY008716	08_BC	CN	1997
AY008717	08_BC	CN	1997
AY727526	31_BC	BR	2004
AY727527	31_BC	BR	2004
AY967803	BC	CN	1996
AY967804	BC	CN	1996
AY967805	BC	CN	1998
AY967807	BC	CN	1996
AY968312	BC	AR	2001
DQ230841	07_BC	TW	2004
EF091932	31_BC	BR	2002
EF368370	07_BC	CN	2005
EF368371	07_BC	CN	2005
EF368372	07_BC	CN	2005
EF420986	07_BC	CN	2005
EU000506	BC	IN	2002
EU000507	BC	IN	2002
EU000508	BC	IN	2002
EU000509	BC	IN	2002
EU000510	BC	IN	2002
EU000511	BC	IN	2002
EU000512	BC	IN	2002
EU000513	BC	IN	2002
EU000514	BC	IN	2002
EU000515	BC	IN	1999
EU000516	BC	IN	1999
EU448296	60_BC	FR	2006

EU839599	BC	DO	2005
FJ388961	BC	CY	2005
GQ365649	BC	BR	2002
GQ365650	BC	BR	2005
GQ365651	BC	BR	2001
GQ365652	BC	BR	2001
GU362013	BC	CN	2008
HM067748	08_BC	CN	2006
HM100716	BC	BR	1992
HM776938	BC	CN	2007
HM776939	57_BC	CN	2007
HQ215552	07_BC	CN	2008
JF719819	08_BC	CN	2007
JN235957	BC	UY	2002
JN692442	60_BC	BR	2002
JN692478	BC	BR	2006
JQ423923	07_BC	CN	2007
JX392378	07_BC	CN	2006
JX392379	07_BC	CN	2006
JX392380	07_BC	CN	2006
JX392381	07_BC	CN	2006
JX392382	07_BC	CN	2006
JX392383	07_BC	CN	2006
JX392384	07_BC	CN	2006
JX679207	57_BC	CN	2009
JX960600	07_BC	CN	2009
JX960601	07_BC	CN	2008
JX960602	07_BC	CN	2009
KC492737	07_BC	CN	2005
KC492738	07_BC	CN	2005
KC503852	07_BC	CN	2005
KC503853	07_BC	CN	2005
KC503854	07_BC	CN	2005
KC503855	07_BC	CN	2005
KC596068	BC	BR	2007
KC870031	64_BC	CN	2009
KC870032	64_BC	CN	2010
KC870033	64_BC	CN	2010
KC870034	62_BC	CN	2010
KC870035	62_BC	CN	2010
KC870036	64_BC	CN	2010
KC870037	62_BC	CN	2010
KC870040	64_BC	CN	2010
KC870042	64_BC	CN	2009

KC870043	64_BC	CN	2009
KC870044	57_BC	CN	2009
KC898975	88_BC	CN	2005
KC898976	BC	CN	2005
KC898977	BC	CN	2005
KC898978	BC	CN	2005
KC898979	88_BC	CN	2005
KC898981	BC	CN	2006
KC898982	BC	CN	2007
KC898983	BC	CN	2007
KC898984	BC	CN	2009
KC898985	57_BC	CN	2009
KC898986	BC	CN	2009
KC898987	110_BC	CN	2009
KC898988	BC	CN	2009
KC898989	BC	CN	2009
KC898990	BC	CN	2009
KC898991	57_BC	CN	2009
KC898993	BC	CN	2009
KC898994	64_BC	CN	2009
KC899006	BC	CN	2005
KC899007	BC	CN	2007
KC899008	57_BC	CN	2009
KC899009	64_BC	CN	2009
KC899010	BC	CN	2009
KC899013	BC	CN	2009
KC899014	BC	CN	2009
KC899015	BC	CN	2009
KC899079	60_BC	IT	2011
KC899080	60_BC	IT	2011
KC899081	60_BC	IT	2011
KC914396	08_BC	CN	2000
KC990124	61_BC	CN	2010
KC990125	61_BC	CN	2007
KC990126	61_BC	CN	2007
KF011493	BC	CN	2010
KF234628	07_BC	TW	2013
KF250366	07_BC	CN	2007
KF250367	61_BC	CN	2007
KF250368	08_BC	CN	2007
KF250369	08_BC	CN	2007
KF250371	07_BC	CN	2007
KF250372	07_BC	CN	2007
KF250373	07_BC	CN	2007

KF250374	07_BC	CN	2007
KF250375	07_BC	CN	2007
KF250376	07_BC	CN	2006
KF250377	07_BC	CN	2007
KF250379	BC	CN	1996
KF250380	BC	CN	1996
KF250381	BC	CN	2007
KF250382	BC	CN	2007
KF250383	BC	CN	2007
KF250384	BC	CN	2007
KF250385	07_BC	CN	2007
KF250397	BC	CN	2010
KF250399	BC	CN	2010
KF250400	57_BC	CN	2010
KF250402	88_BC	CN	2009
KF250405	BC	CN	2012
KF250406	BC	CN	2012
KF250409	118_BC	CN	2012
KF250410	BC	CN	2012
KF312642	08_BC	CN	2007
KF835493	08_BC	CN	2007
KF835494	07_BC	CN	2007
KF835495	08_BC	CN	2007
KF835496	08_BC	CN	2007
KF835497	08_BC	CN	2007
KF835500	08_BC	CN	2007
KF835501	08_BC	CN	2007
KF835506	08_BC	CN	2007
KF835507	08_BC	CN	2007
KF835508	08_BC	CN	2007
KF835509	08_BC	CN	2007
KF835510	07_BC	CN	2007
KF835511	07_BC	CN	2007
KF835512	08_BC	CN	2007
KF835521	BC	CN	2007
KF835524	110_BC	CN	2007
KF835525	BC	CN	2007
KF835526	BC	CN	2007
KF835529	08_BC	CN	2007
KF835530	08_BC	CN	2007
KF835531	08_BC	CN	2007
KF835532	07_BC	CN	2007
KF835534	08_BC	CN	2007
KF835535	08_BC	CN	2007

KF835539	BC	CN	2007
KF835540	BC	CN	2007
KF835541	08_BC	CN	2007
KF835544	BC	CN	2007
KF835545	08_BC	CN	2007
KF835547	08_BC	CN	2007
KJ849795	BC	BR	2010
KT183073	BC	ZA	2009
KT183074	BC	ZA	2009
KT183075	BC	ZA	2009
KT183076	BC	ZA	2009
KT183077	BC	ZA	2009
KT427760	BC	BR	2010
KT427778	BC	BR	2010
KT427790	BC	BR	2010
KT427822	BC	BR	2010
KT960983	BC	CN	2014
KU886698	BC	CN	2014
KU992928	85_BC	CN	2014
KU992929	85_BC	CN	2014
KU992930	85_BC	CN	2014
KU992931	85_BC	CN	2014
KU992932	85_BC	CN	2014
KU992934	85_BC	CN	2014
KU992935	85_BC	CN	2014
KU992936	85_BC	CN	2014
KU992937	85_BC	CN	2014
KX582249	86_BC	CN	2016
KX58225	86_BC	CN	2016
KX582251	86_BC	CN	2016
KY406739	BC	CN	2014
MF109555	07_BC	GB	2014
MF109595	BC	GB	2014
MF373189	BC	SE	2013
MG742702	BC	CN	2015
MK258683	85_BC	CN	2014
MK258684	85_BC	CN	2012
MK792287	BC	CN	2015
MK792288	BC	CN	2018
MK792289	BC	CN	2018
MN172222	108_BC	ES	2016
MN172223	108_BC	ES	2018
MN172224	108_BC	ES	2018
MN172225	108_BC	ES	2017



MN659380	BC	CN	2018
MN659381	BC	CN	2018
MT021910	07_BC	KR	2016
MT559132	108_BC	ES	2017
MT635156	BC	VN	2010
MT635171	07_BC	VN	2013
MW145181	BC	CN	2018
MW419275	110_BC	CN	2012

Legend: AR - Argentina, BR - Brazil, CN - China, CY - Cyprus, DO - Dominican Republic, ES - Spain, FR - France, IN - India, IT - Italy, KR - Korean Republic, MM - Myanmar, SE - Sweden, TW - Taiwan, UK - United Kingdom, UY - Uruguay, VN - Vietnam and ZA - South Africa

## Capítulo VI

---

Predição de características moleculares no subtipo C do HIV-1 encontrados no Brasil e em várias partes do mundo.

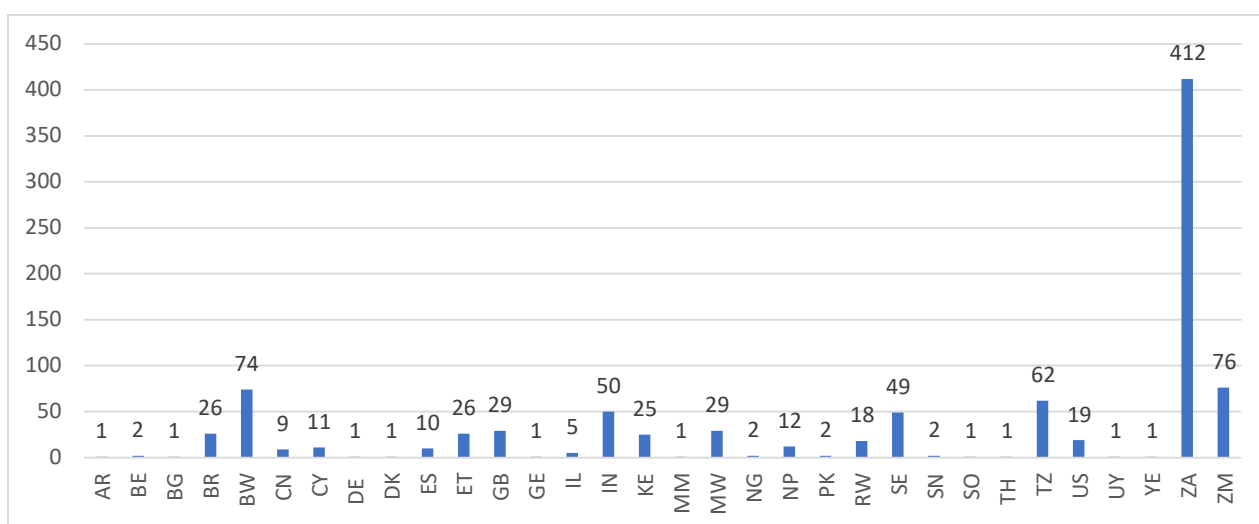
## Predição de características moleculares no subtipo C do HIV-1 encontrados no Brasil e em várias partes do mundo.

### 1- Resultados e discussão

Foram coletadas 2106 amostras de genomas com comprimento quase total (NFLG - *Near Full Length Genome*) do banco de dados de Los Alamos (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) pertencentes ao subtipo C do HIV-1 de 33 países, incluindo 25 sequências brasileiras, baseado na aplicação dos seguintes filtros: subtipo C e genoma completo. Adicionamos a amostra 011 da nossa coorte, somando assim, 26 amostras do Brasil e um total de 2107 sequências NFLG. Foram excluídas 10 amostras que não apresentavam informação de ano de obtenção e uma amostra sem a informação do local de origem. Foi observada uma grande discrepância no que se refere à origem das sequências do subtipo C, com os países africanos representando 85% de todas as sequências NFLG do subtipo C disponíveis até 16 de fevereiro de 2022: 42,4% (N=894) provenientes da África do Sul, 18,9% (N=400) da Zâmbia, 10,6% (N= 225) da Botswana, 6,7% (N=243) do Malawi e 6,1% (N= 130) da Tanzânia. Assim, foi aplicada a ferramenta CD-HIT EST (HUANG et al., 2010) nos conjuntos de dados com mais de 100 amostras/país (Botswana, Malawi, Tanzânia, África do Sul e Zâmbia), para isso foram usados os seguintes parâmetros: *query*: consenso do subtipo C disponível nos Los Alamos, 0,95% de identidade e o filtro de sequências mantendo o comprimento mínimo de 7000 pares de base de extensão. Após o uso do CD-HIT EST (HUANG et al., 2010) foram disponibilizadas 960 amostras ao total, o que corresponde 45,5% (1147 sequências a menos) do conjunto de dados original. Os 33 países representados foram mantidos juntamente com a mesma amostragem anual: 33 anos sem representantes dos anos de 1987, 1988 e 2018. O novo conjunto de amostras contendo 960 sequências (Gráfico 1) foi alinhado pela plataforma on-line MAFFT (<https://mafft.cbrc.jp/alignment/software/>) e editado manualmente pelo programa Bioedit (HALL, 1999). Utilizamos as sequências referências do subtipo C do Los Alamos, a sequência referência da HXB2 (por gene) e a consenso C (genoma quase completo), disponíveis em <https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>, para o alinhamento e edição. Os conjuntos de dados foram alinhados, editados e cortados de acordo com as posições gênicas da HXB2 e da sequência consenso C (<https://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/HXB2.html>). Foram obtidos 10

conjuntos de dados: (GAG, POL, VIF, VPR, NEF, REV, VPU, ENV, NEF e LTR). Após essa etapa, as sequências nucleotídicas foram ordenadas por ano e avaliadas no Bioedit (HALL, 1999) para as análises de aminoácidos e motivos estruturais.

Gráfico 1 – Amostragem em números absolutos das 960 sequências nucleotídicas do subtipo C do HIV-1 por país. Legenda: AR- Argentina; BE- Bélgica; BG- Bulgária; BR- Brasil; BW- Botsuana; CN- China; CY- Chipre; DE- Alemanha; DK- Dinamarca; ES- Espanha; ET- Etiópia, GB- Reino Unido; GE- Geórgia; IL- Israel; KE- Quênia; MM- Mianmar; MW- Malawi; NG- Nigéria; NP- Nepal; PK- Paquistão; RW- Ruanda; SE- Suécia; SN- Senegal; SO- Somália; TH- Tailândia; US- Estados Unidos; UY- Uruguai; YE- Iémen; ZA- África do Sul; ZM- Zâmbia.



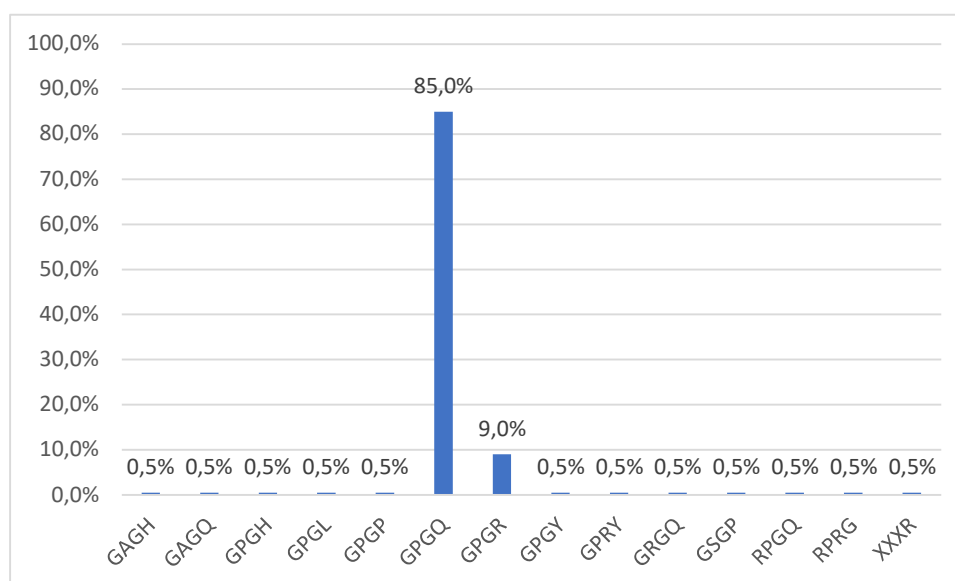
Fonte: autoria própria

### 1.1 Análise da utilização dos receptores de quimiocinas pelos isolados do subtipo C, com base na sequência de aminoácidos.

A entrada do HIV-1 nas células-alvo é um processo que se inicia através da adsorção da gp120 viral com a proteína CD4 do hospedeiro, o que resulta em alterações conformacionais para permitir a participação de um correceptor para o vírus na célula humana. Entre estes, já foram descritos o CCR2, CCR3, CCR5, CCR8, CXCR4, CXCR6, CX3CR1, CHENR23 ou RDC1, embora CCR5 e CXCR4 constituam os correceptores mais frequentemente utilizados *in vivo* pelo vírus (ARENZANA-SEISDEDOS et al., 2006; PIERSON & DOMS, 2003). Alguns autores já mostraram a associação entre a preferência de uso do correceptor e a velocidade de progressão para a doença (CHENG-MAYER et al., 1998; HO, et al., 2005). Os mecanismos envolvidos na categorização em lento-progressores (R5) ou progressores (X4) na infecção, ocorreram baseado em clones celulares que expressam estas moléculas em suas superfícies e

pode-se ser constatado pela a depleção da contagem de linfócitos T CD4+ desencadeada por vírus X4-trópicos (CHENG-MAYER et al., 1998; HO, et al., 2005). Com base nas sequências de nucleotídeos aplicadas à ferramenta on-line Geno2Pheno (<https://coreceptor.geno2pheno.org/>), das 960 amostras virais do subtipo C utilizadas nesta pesquisa (100%), 760 amostras (79,2%) apresentaram preferência na utilização do CCR5, enquanto as 200 amostras restantes (20,8%) apresentaram preferência no uso do correceptor CXCR4. Avaliando as sequências de aminoácidos das 760 amostras CCR5, o tetrapeptídeo GPGQ apresentou-se conservado na região da alça V3 na proteína gp120, enquanto as 200 amostras com preferência para o correceptor CXCR4, apresentaram uma diversidade maior neste motivo (Gráfico 2).

Gráfico 2 – Diversidade de tetrapeptídeos na alça V3 da proteína gp120 das amostras com tropismo CXCR4.



Fonte: autoria própria

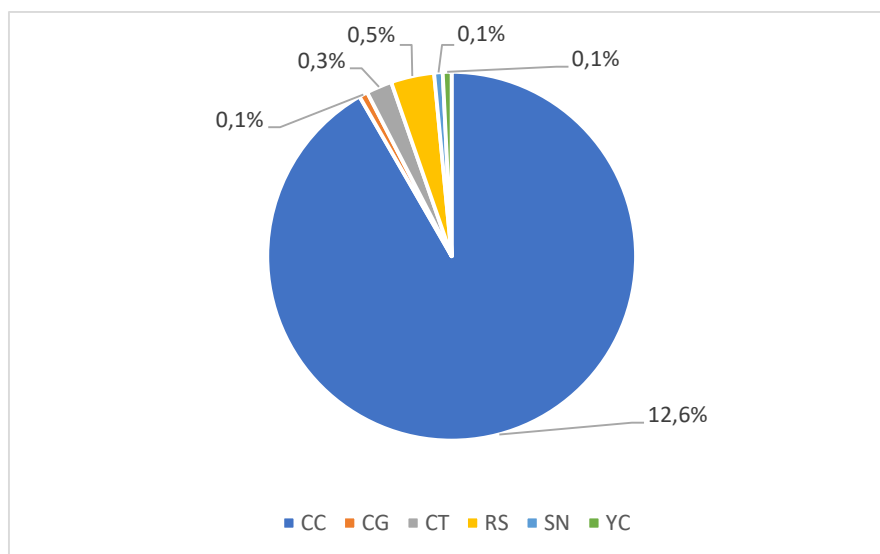
Das 26 sequências do subtipo C do HIV-1 representantes do Brasil, apenas uma amostra, identificada no estado de São Paulo no ano de 2002, revelou o tetrapeptídeo GPGR. As demais amostras brasileiras corroboram com os dados da literatura (SUCUPIRA et al., 2012), os quais apontam que 83% das cepas de HIV-1 no país têm tropismo para o correceptor CCR5 (R5-trópicos) e a nossa amostra isolada em 2019 na região Nordeste, não pode-se estimar tropismo e tetrapeptídeo devido às sobreposições de picos no eletroferograma da amostra 011.

Por outro lado, apenas 17% das cepas brasileiras apresentaram o tropismo para CXCR4 (X4), com motivo GPGR (SUCUPIRA et al., 2012). Considerando apenas o genótipo C, de acordo com o trabalho de Kalu e colaboradores (2017), associado ao dado de ancestralidade da cepa C brasileira, que é proveniente do centro-leste do continente africano (BELLO et al., 2008) nestes países os vírus são quase exclusivamente R5-tropicos (KALU et al., 2017).

## 1.2 Busca pelo dipeptídeo di-cisteína ou cisteína-serina nas posições nas 30 e 31.

De acordo com estudos prévios (SATISCHANDRA et al., 2000; WADIA et al., 2001) 90% das cepas do subtipo C do HIV-1 apresentam o dipeptídeo C-S (cisteína-serina) nas posições 30 e 31 respectivamente na proteína Tat. Quando o motivo dicisteína (C30-C31) está presente, parece estar associado à uma capacidade de infecção de células do sistema nervoso central, os astrócitos, e durante essa infecção, favorece a expressão do gene que codifica da proteína MCP-1 (Proteína Quimiotática de Monócitos-1), que induz o recrutamento destas células, permitindo o reconhecimento da infecção no local (sistema nervoso central) e conduz à resposta inflamatória, condicionando o surgimento do quadro clínico de demência associada ao HIV (RANGA et al., 2004). Avaliando as posições 30 e 31 na proteína Tat do nosso conjunto de dados (N=960 [100%]), 828 sequências (86,2%) revelaram o dipeptídeo C-S e apenas 132 amostras (13,8%) encontradas na África do Sul, Alemanha, Bélgica, Brasil, Botsuana, China, Chipre, Espanha, Etiópia, Iêmen, Índia, Malawi, Nepal, Reino Unido, Suécia, Tailândia, Tanzânia e Zâmbia apresentaram substituições 30C-31X (CC, CG, CT, RS, SN, e YC) nestas posições (Gráfico 3).

Gráfico 3 – Frequência relativa de substituições no dipeptídeo C-X encontrados entre as 132 de 960 amostras do subtipo C do Brasil e do mundo. Legenda: C: cisteína; G: glicina; T: treonina; R: arginina; S: serina; N: asparagina; Y: tirosina.



Fonte: autoria própria

Considerando apenas as amostras brasileiras (N=26 [100%]) isoladas entre o período de 1992 a 2019, a substituição C-C foi constatada em seis amostras (23%) ao longo do intervalo de 27 anos: a primeira isolada em 2002 no estado de São Paulo e as demais em: 2009 (1 - São Paulo), 2010 (2 – Pernambuco e São Paulo), 2017 (1 – Rio Grande do Sul) e 2019 (1), sendo esta última identificada no estado da Bahia. Não foram encontrados dados clínicos da amostra 011 que pudessem fornecer robustez a esta inferência. De acordo com o estudo de Ranga e colaboradores (2004), a substituição da cisteína-31 na proteína TAT têm uma profunda consequência negativa nas propriedades biológicas dos vírus, a perda da capacidade quimiotática direcionada à monócitos no sistema nervoso central. A avaliação do conjunto de dados das amostras brasileiras (N=26) no servidor do DATAMONKEY revelou que a posição 31 em Tat não sofre pressão seletiva de nenhum tipo.

### 1.3 Avaliar a presença dos motivos GSXXS e WL na proteína *nef*.

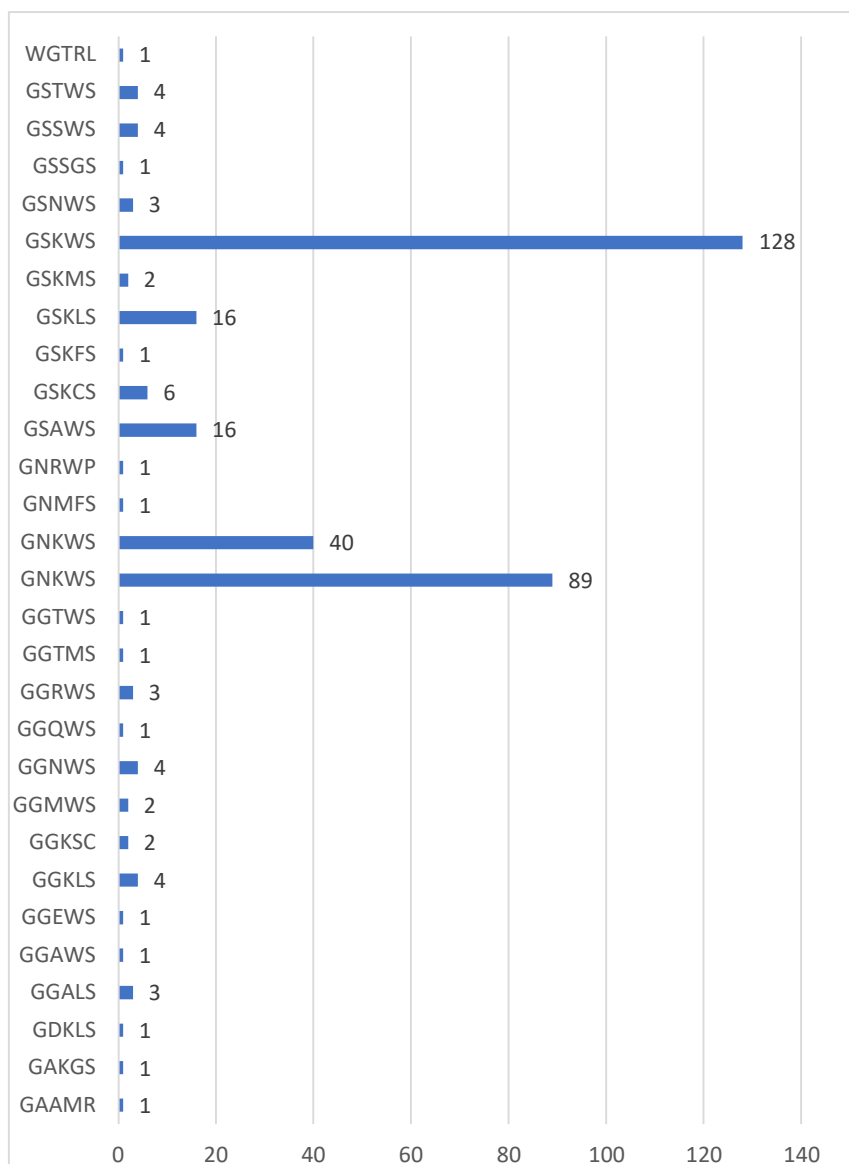
O gene *nef* codifica a proteína NEF que participa diretamente da regulação do ciclo viral, e é fortemente traduzida durante as etapas iniciais deste ciclo (PEREIRA et al., 2016). Esta proteína está envolvida em vários mecanismos intracelulares como, o transporte de vesículas pelo sistema de endomembranas, mecanismos de evasão contra a resposta imune, internalizando o complexo de histocompatibilidade de classe I e da liberação de novos vírions, pelo efeito de miristoilção e brotamento através da membrana (PEREIRA et al., 2016; USAMI et al., 2015). Algumas características moleculares nesta proteína foram

relacionadas com a progressão da patologia (SCRIBA et al., 2002). De acordo com Scriba e colaboradores (2002), o motivo GGXXS é altamente conservado no genótipo C (posições 2-6 na proteína NEF) e mutações não-sinônimas ou deleções que reduzam o número de resíduos de glicina (ligado ao ácido mistoleico), podem induzir a progressão lenta da doença, por interferir nos eventos de miristoilação.

Em nosso trabalho 65% (N=621) das amostras mantinham este motivo estrutural (GGKWS), enquanto amostras de 22 países (África do Sul, Argentina, Brasil, Bélgica, Botsuana, China, Chipre, Dinamarca, Espanha, Estados Unidos, Etiópia, Índia, Israel, Malawi, Nepal, Quênia, Reino Unido, Ruanda, Suécia, Tanzânia, Uruguai e Zâmbia), correspondendo a 35% (N=339) do conjunto de dados apresentaram 29 variações na sequência de aminoácidos (Tabela 1). Avaliando apenas as 26 sequências nucleotídicas identificadas no Brasil, três motivos foram identificados (GNKWS [N= 23], GSKWS [N= 2] e GSKLS [N= 1]) e todos com menor expressão de resíduos de glicinas na proteína. A análise de pressão seletiva entre as 26 amostras brasileiras revelou que há apenas seleção negativa para o resíduo de Lisina (K) na posição 4 da proteína e entre os 29 motivos variantes, apenas 10 destes mantêm a lisina nesta posição. Outra assinatura WL (posições 56-57) corresponde a um sítio de clivagem proteolítica, e de acordo com Scriba e colaboradores (2002), é altamente conservado na cepa C. Nossos resultados corroboram com este dado, quando nossa análise revelou que apenas 8,9% (N= 85) das amostras apresentaram substituições nestas posições e nenhuma das 85 sequências eram brasileiras.



Gráfico 4 – Frequência absoluta das variações no motivo GGXXS da proteína NEF de 339 amostras do conjunto de dados.



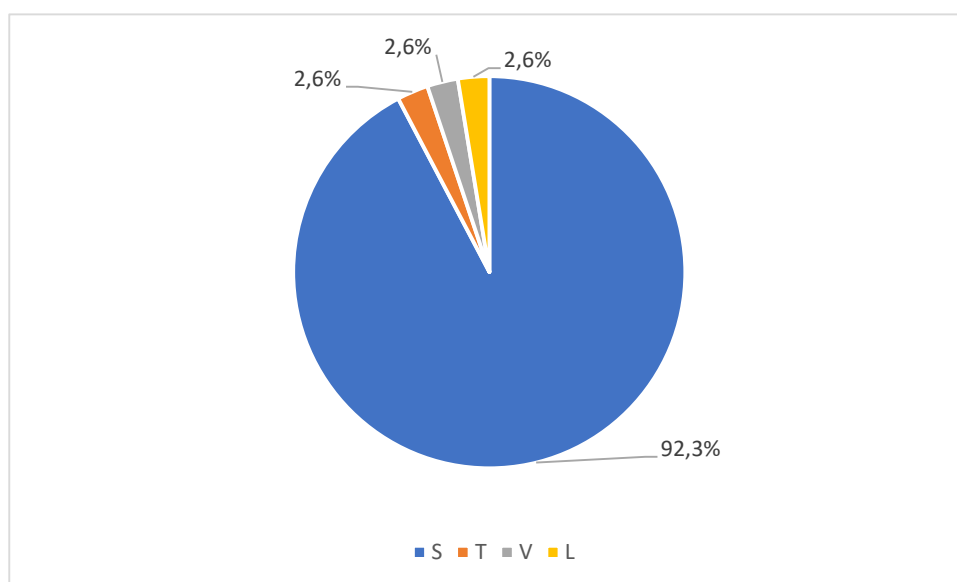
Fonte: autoria própria.

#### 1.4 Verificar a presença do motivo 483L-484Y, mutações ou polimorfismos na região GAG-protease

Os produtos gênicos da região *gag-protease* são diretamente relacionados com marcadores de progressão da doença (FREED et al., 2007) e é descrito na literatura que as substituições nas posições 67, 483 e 484 estão fortemente associadas à capacidade replicativa (BROCKMAN et al., 2010; KIGUOYA et al., 2017; WRIGHT et al., 2010). Na proteína GAG, a nível da matriz viral, foi descrita uma associação entre a posição 67A e a diminuição

da capacidade replicativa no subtipo C, que difere de outros genótipos virais, como o subtipo B que mantém uma serina (S) nesta posição, relacionado a uma maior capacidade replicativa. Em nosso conjunto de dados, a alanina (A) apresentou-se altamente conservada nesta posição, apenas 39/960 (4%) amostras apresentam outros aminoácidos (Gráfico 5). Avaliando as localidades das 39 sequências, foi observado que nenhuma delas era do Brasil (África do Sul, Botsuana, Chipre, Etiópia, Índia, Reino Unido, Nepal, Ruanda, Suécia, Tanzânia, e Zâmbia).

Gráfico 5 – Frequência relativa das substituições de aminoácidos na posição 67X em 39 (100%) cepas do subtipo C. Legenda: L – leucina; S – serina; T – treonina e V – valina.



Fonte: autoria própria.

De acordo com os trabalhos de Brockman e colaboradores (2010), e Kiguoya e colaboradores (2017), a assinatura LY nas posições 483-484 na proteína GAG (p6) é um determinante de maior capacidade replicativa viral, presente no subtipo B e alguns recombinantes do HIV-1, porém, no genótipo C a deleção destas posições é mais comum. Em contrapartida, quando presente no subtipo C capacidade replicativa é aumentada. Em nossa pesquisa 97,2% (N= 934) das sequências apresentaram a deleção (-/-) nas posições 483 e 484, enquanto apenas 2,8% (N=26) apresentaram variações nestas posições (Tabela 1) e nenhuma das 26 amostras revelaram a assinatura 483L-484Y. Avaliando apenas as amostras do Brasil (N= 26), apenas 3 apresentaram informação biológica nestas posições: KT427817, identificada no estado de São Paulo (região Sudeste); KT427736, no estado de

Pernambuco (região Nordeste) e KU749391, sem informação de estado ou região. Nos três casos não há descrição da assinatura 483L-484Y.

Tabela 1 – Descrição da informação biológica das 26 sequências nas posições 483 e 484 na proteína GAG. Legenda: BR- Brasil; BW- Botsuana; ET- Etiópia; IL- Israel; MW- Malawi; RW- Ruanda; SE- Suécia e ZM- Zâmbia.

<b>Ano</b>	<b>País</b>	<b>483</b>	<b>484</b>
		<b>L</b>	<b>Y</b>
<b>1996</b>	<b>BW</b>	~	P
<b>1998</b>	<b>IL</b>	P	~
<b>1999</b>	<b>IL</b>	P	~
<b>1999</b>	<b>IL</b>	P	~
<b>2004</b>	<b>SE</b>	~	C
<b>2007</b>	<b>ZM</b>	T	~
<b>2007</b>	<b>MW</b>	~	P
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>RW</b>	P	~
<b>2008</b>	<b>SE</b>	P	~
<b>2008</b>	<b>ET</b>	~	H
<b>2008</b>	<b>ET</b>	P	~
<b>2009</b>	<b>SE</b>	P	~
<b>2009</b>	<b>SE</b>	I	~
<b>2009</b>	<b>ZM</b>	P	~
<b>2009</b>	<b>BR</b>	H	~
<b>2010</b>	<b>BR</b>	H	~
<b>2010</b>	<b>BR</b>	S	~
<b>2011</b>	<b>SE</b>	P	~
<b>2013</b>	<b>SE</b>	P	~

Fonte: autoria própria.

A função Gag-protease pode afetar significativamente a taxa de progressão da doença, refletindo na capacidade global de replicação de isolados que difere entre os subtipos, permitindo classificar as cepas em não-progressores e progressores de longo prazo, indicando que esse parâmetro influencia significativamente a taxa de progressão da doença (KIGUOYA et al., 2017). Avaliando algumas posições-chave na função da proteína nossos dados corroboram com demais autores (BROCKMAN et al., 2010; KIGUOYA et al., 2017; WRIGHT et al., 2010), quando trata da menor capacidade de replicação pelo subtipo C e análises moleculares relacionam estes achados à uma expansão do subtipo C, o que justifica os dados de prevalência a nível mundial, 46,6% das infecções por este subtipo (HEMELAAR et al., 2019).

Embora nossa pesquisa não avalie o potencial biológico das substituições descritas, os resultados das análises *in-silico* apontam que a maioria dos vírus pertencentes ao subtipo C circulantes no mundo possuem características moleculares semelhantes e quando delimitamos as observações por país, as amostras identificadas no Brasil apresentam os mesmos achados. Embora seja preciso interpretar os dados descritos acima com cautela, uma vez que o número de sequências analisadas foi baixo (N= 26), associado a isto, não foram identificados dados clínicos da nossa amostra 011 e nem das amostras resgatadas do banco de dados. De acordo com Graf e colaboradores (2021), Kiguoya e colaboradores (2017), e Souto e colaboradores (2021), as propriedades biológicas do subtipo C parecem favorecer sua própria transmissão, a baixa capacidade replicativa, garante maior tempo de sobrevivência das células T CD4+ infectadas que continuam liberando novos vírions, resultando em maior oportunidade de transmissão, justificando o aumento de prevalência detectados em alguns países da África oriental (HEMELAAR et al., 2019; KIGUOYA et al., 2017) e no Brasil (ARRUDA et al., 2018; GRAF et al., 2021), por exemplo. Um dado que corrobora com estas observações foi apresentado no trabalho de Silveira e colaboradores (2017), que constataram uma maior contagem de linfócitos T CD4+ em pessoas que vivem com HIV/AIDS com genótipo C no país.

## 2- Conclusão

Considerando as características associadas a dispersão e transmissibilidade levantadas nesta busca apenas 26 sequências do subtipo C do HIV-1 foram identificadas no Brasil e foi possível verificar que um número pequeno de amostras apresentam algumas assinaturas de

aminoácidos que já foram descritas em mecanismo patogênicos. Outro limite foi a ausência de informações clínicas disponíveis no banco de dados e da nossa amostras 011, que poderiam justificar alguns resultados desta pesquisa. Neste trabalho foram apresentadas outras assinaturas de aminoácidos em proteínas virais que ainda tem funções desconhecidas, portanto, são necessários mais estudos para investigar o potencial biológico destas substituições em razão da patogenicidade e transmissibilidade do subtipo C no país.

### 3- Referências

ARENZANA-SEISDEDOS F, PARMENTIER M: Genetics of resistance to HIV infection: Role of co-receptors and co-receptor ligands. *Semin Immunol* 2006, 18(6):387-403. <https://doi.org/10.1016/j.smim.2006.07.007>

ARRUDA, M. B., BOULLOSA, L. T., CARDOSO, C. C., DA COSTA, C. M., BRITES, C., DE LIMA, S. T. KAMINSKI, H. T., ALEIXO, A. W., ESPOSITO, O. P., CAVALCANTI, A. M., RIEDEL, M., COUTO-FERNANDEZ, J. C., FERREIRA, S. B., DE OLIVEIRA, I. C., PORTAL, L. E., WOLF, H. H., FERNANDES, S. B., PARDINI, M. I M. C., FEITEIRO, M. V., TOLENTINO, F. M., DIAZ, R. S., LOPES, G. I., FRANCISCO, R. B., VÉRAS, N. M., PIRES, A. F., FRANCHINI, M., MESQUITA, F., TANURI, A., HIV-BResNet. Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): a survey of treatment-naive individuals. *Journal of the International AIDS Society*, 21(3), e25032. 2018

BROCKMAN, M., BRUMME, Z., BRUMME, C., MIURA, T., SELA, J., ROSATO, P., KADIE, C., CARLSON, J., MARKLE, T., STREECK, H., KELLEHER, A., MARKOWITZ, M., JESSEN, H., ROSENBERG, E., ALTFELD, M., HARRIGAN, P., HECKERMAN, D., WALKER, B., ALLEN, T. Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. *J Virol* 84:11937–11949. 2010. <https://doi.org/10.1128/JVI.01086-10>

CHENG-MAYER, C., SETO, D., TATENO, M., LEVY, J., A. Biologic features of HIV-1 that correlate with virulence in the host. *Science* 1988, 240(4848):80-82

FREED EO & MARTIN MA. HIVs and their replication, p 2107–2185. *In* Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (ed), *Fields virology*, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA. 2007.

GRÄF, T., BELLO, G., ANDRADE, P., ARANTES, I., PEREIRA, J. M., DA SILVA, A. B. P., VEIGA, R. V., MARIANI, D., BOULLOSA L, T., ARRUDA, M, B., FERNANDEZ, J, C, C., DENNIS, A, M., RASMUSSEN D, A., TANURI, A. (2021). HIV-1 molecular diversity in Brazil unveiled by 10 years of sampling by the national genotyping network. *Scientific Reports*, 11(1). doi:10.1038/s41598-021-94542-5

HALL, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 1999

- HEMELAAR, J., LOGANATHAN, S., ELANGO VAN, R. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infectious Disease*. 19 (2): 143-155 2019. [https://doi.org/10.1016/S1473-3099\(18\)30647-9](https://doi.org/10.1016/S1473-3099(18)30647-9)
- HUANG, Y., NIU, B., GAO, Y., FU, L., LI, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, n5, pp. 680-682. 2010
- HO, S.H., SHEK, L., GETTIE, A., BLANCHARD, J., CHENG-MAYER, C. V3 loop determined coreceptor preference dictates the dynamics of CD4+T-cell loss in simian-human immunodeficiency virus infected macaques. *J Virol* 2005, 79(19):12296-12303. <https://doi.org/10.1128/JVI.79.19.12296-12303>
- KALU, A., W., TELELE, N., F., GEBRESELASIE, S., FEKADE, D., ABDURAHMAN, S., MARRONE, G., SONNERBORG, A. Monophyletic HIV-1C epidemic in Ethiopia is dominated by CCR5-tropic viruses-an analysis of a prospective country-wide cohort. *BMC Infectious Diseases*, 17:37. 2017.
- KIGUOYA, M.W., MANN, J.K., CHOPERA, D., GOUNDER, K., LEE, G.Q., HUNT, P.W., MARTIN, J. N., BALL, B., KIMANI, J., BRUMME, Z. L., BROCKMAN, M.A., NDUNG'U, T. Subtype Specific Differences in Gag-Protease-Driven Replication Capacity Are Consistent with Intersubtype Differences in HIV-1 Disease Progression. *Journal of Virology*; 91(13) e00253-17. 2017.
- PEREIRA, E. A., DA SILVA, L. P. HIV-1 Nef: Taking Control of Protein Trafficking. *Traffic*. 17; 9: 976-996. <https://doi.org/10.1111/tra.12412>.
- PIERSON TC, DOMS RW. HIV-1 entry and its inhibition. *Curr Top Microbiol Immunol* 2003, 281:1-27. [https://doi.org/10.1007/978-3-642-19012-4\\_1](https://doi.org/10.1007/978-3-642-19012-4_1).
- RANGA, U., SHANKARAPPA, R., SIDDAPPA, N. B., RAMAKRISHNA, L., NAGENDRAN, R., MAHALINGAM, M., MAHADEVAN, A., JAYASURY, N. Tat Protein of Human Immunodeficiency Virus Type 1 Subtype C Strains Is a Defective Chemokine. *American Society for Microbiology. Journal of Virology*, Mar. 2004, p. 2586–2590 Vol. 78, No. 5. <https://doi.org/10.1128/JVI.78.5.2586-2590.2004>
- SATISHCHANDRA, P., A. NALINI, M. GOURIE-DEVI, N. KHANNA, V. SANTOSH, V. RAVI, A. DESAI, A. CHANDRAMUKI, P. N. JAYAKUMAR, AND S. K. SHANKAR. Profile of neurologic disorders associated with HIV/AIDS from Bangalore, south India. *Indian J. Med. Res.* 111:14-2. 2000.
- SCRIBA, T. J., VILLIERS, T., TREURNICHT, F. K., MEGEDE, J. Z., BARNETT, S. W., ENGELBRECHT, S., RESENBURG, E. J. V. Characterization of the South African HIV type 1 Subtype C complete 5' Long Terminal Repeat nef and regulatory genes. *AIDS Research and Human Retroviruses*. V. 18. N.2. PP 149-159. 2002.
- SOUTO, B., TRIUNFANTE, V., SANTOS-PEREIRA, A., MARTINS, J., ARAÚJO, P. M. M., OSÓRIO, N. S. Evolutionary dynamics of HIV-1 subtype C in Brazil. *Sci Rep.*; 11: 23060. 2021.

SUCUPIRA, M. C. A., SANABANI, S., CORTES, R. M., GIRET, M. T. M., TOMIYAMA, H., SAUER, M. M., SABINO, E. C., JANINI, L. M., KALLAS, E. G., DIAZ, R. S. Faster HIV-1 disease progression among Brazilian individuals recently infected with cxcr4-utilizing strains. *PlosOne*, 2012; 7: e30292. <https://doi.org/10.1371/journal.pone.0030292>

USAMI, Y., WU, Y., GOTTLINGER, H.G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*. 526 (7572): 218–223. <https://doi.org/10.1038/nature15400>. 2015.

WADIA, R., PUJARI, S., KOTHARI, S., UDHAR, M., KULKARNI S., BHAGAT, S., NANIVADEKAR, A. Neurological manifestations of HIV disease. *Journal of Associations Physicians India*. 49: 343-348. 2001.

WRIGHT, J. K., BRUMME, Z. L., CARLSON, J. M., HECKERMAN, D., KADIE, C. M., BRUMME, C. J., WANG, B., LOSINA, E., MIURA, T., CHONCO, F., VAN DER STOK, M., MNCUBE, Z., BISHOP, K., GOULDER, P. J. R., WALKER, B.D., BROCKMAN, M.A., NDUNG'U, T. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. *J Virol* 84:10820–10831. 2010. <https://doi.org/10.1128/JVI.01084-10>

## 5 DISCUSSÃO

Os estudos de epidemiologia molecular com o HIV-1 no Brasil, descrevem que a cepa B é a prevalente, seguido do subtipo C e subtipo F1. Além das formas genéticas puras, as cepas recombinantes BF1 e BC também são detectadas (ARRUDA et al., 2018 e GRAF et al., 2021). Um estudo de caracterização genética e molecular acerca dos recombinantes BF1, destacou as propriedades virais de amostras locais e mundias (SOUZA et al., 2019), desta vez utilizamos as ferramentas de Biologia Molecular e Bioinformática para estudar as formas não-B (F1, C e BC) do HIV-1 no Brasil, até a região Nordeste.

As análises filogenética e filodinâmica com os novos isolados F1 da nossa coorte revelam que as variantes brasileiras F1 do HIV-1 tem uma origem única. Este dado pôde ser constatado devido à topologia da nossa árvore contendo mais de 670 sequências F1 de vários países. Todas as amostras identificadas no Brasil, agruparam no mesmo grupo monofilético (clado F1I) e estimamos o tempo do ancestral comum mais recente deste subsubtipo em 1974 (SILVA et al., 2022). De fato, duas grandes linhagens F1 já foram descritas no mundo (GUIMARÃES et al., 2009), mas em nossa pesquisa foram incluídos fragmentos homólogos BF1 à região F1 do gene *pol* e pode-se verificar que a cepa recombinante BF1 brasileira tem relações filogenéticas robustas (suportadas por *aLRT* e *bootstrap*) com os recombinantes BF1 de outras localidades, sugerindo que o genótipo local contribua com epidemiologia por BF1 em outros países.

Sobre o subtipo C do HIV-1, as análises filogeográficas apontaram que o estado do Paraná, região sul do país, tenha sido o local de chegada deste vírus, constituindo assim a linhagem principal desde genótipo (BELLO et al., 2012). Essa hipótese foi comprovada com os estudos de Delatorre e colaboradores (2013) e Graf e colaboradores (2015), os quais também foram baseados em análises filogeográficas Bayesianas, revelando que a epidemia pelo subtipo C em outras regiões do Brasil, é causada pela cepa encontrada na região Sul. Nestas duas pesquisas foram estimados os períodos que o subtipo C alcançou em cada território: no Sudeste e no Centro-Oeste, no início da década de 80, e na região Norte, no fim dos anos 80, não havendo, portanto, a estimativa para a região Nordeste. Os resultados das nossas análises suportam a hipótese de que a cepa que circula na região Nordeste, de fato, descende da linhagem brasileira principal, e identificamos movimentos migratórios para a região Nordeste a partir das regiões Sul, Sudeste e Centro-Oeste. A estimativa do tempo do ancestral comum mais recente na região Nordeste foi durante meados da década de 80 (1985) (BELLO et al., 2012;



DELATORRE et al., 2013; OLIVEIRA et al., 2021). Embora este genótipo alcançou o Brasil na região Sul, durante a década de 70, múltiplas transmissões já foram constatadas a partir das regiões Sudeste e Norte (BELLO et al., 2012; DELATORRE et al., 2013; OLIVERA et al., 2021). No ano de 2013, o trabalho de Delatorre e colaboradores evidenciou introduções adicionais do subtipo C pelo estado do Rio de Janeiro, e em nossa pesquisa, evidenciamos outra entrada adicional pelo estado de Roraima, norte do país. Avaliando as características moleculares de cada representante (sequência) das introduções adicionais, verificamos que cada uma delas apresenta assinaturas de aminoácidos particulares, não compartilhadas com a sequência consenso da linhagem C principal, dando robustez a nossa hipótese.

Através da identificação da amostra recombinante B/C (223) da nossa coorte, foi realizado um levantamento no banco de dados dos Los Alamos por sequências B/C brasileiras de comprimento quase total. A busca retornou 16 amostras com diferentes padrões de recombinação, sendo que três destas (GQ365652, KT42778 e KT427790), identificadas nas regiões Sul e Sudeste, apresentavam apenas uma inserção pertencente ao subtipo B a nível dos genes *integrase* e *VIF*. As análises de recombinação revelaram que os pontos de recombinação são extremamente semelhantes à nossa 223, além disso em todas as análises filogenéticas baseadas nos fragmentos B e C da amostra 223, as quatro sequências agrupam no mesmo clado com bons suportes estatísticos (*aLRT* e *bootstrap*). De acordo com Robertson (2000), para definir uma CRF, pelo menos três genomas quase completos do HIV-1 devem ser caracterizados a partir de indivíduos epidemiologicamente não ligados, mostrando padrões de mosaico idênticos e agrupados em árvores filogenéticas juntamente dos CRFs previamente definidos. Aqui, identificamos três genomas quase completos e um genoma parcial com os mesmos padrões de recombinação quem agrupam no mesmo clado em quatro topologias diferentes em árvores de máxima verossimilhança, que foram desenvolvidas em dois programas: IQTree (TRIFINOPOULOS et al., 2016) e RDP5 (MARTIN et al., 2021). Avaliando estes resultados juntamente com os dados dos pacientes descritos nas publicações que sequenciaram as três amostras: GQ365652, KT42778 e KT427790 (PASSAES et al., 2009; PESSOA et al., 2016), os autores descreveram que não há vínculo epidemiológico entre KT427778 e KT427790, uma vez que ambos os isolados foram obtidos de dois pacientes do mesmo estado (Rio de Janeiro), enquanto o isolado GQ365652 foi é proveniente de um paciente na região sul do Brasil. Deste modo, aqui descrevemos uma nova CRF\_BC brasileira, encontrada em três regiões: Sul, Sudeste e Nordeste.

A caracterização de todos os genomas BC de comprimento quase total recolhidos dos

Los Alamos (N=225) revela que a China e Brasil detêm o maior número de sequências depositadas neste banco de dados. A circulação de mais de um subtipo puro é uma característica comum a estes dois territórios e de acordo com Han e colaboradores (2013), devido à alta prevalência dos subtipos B e C na China, os eventos de recombinação intersubtipos são favorecidos, impulsionando o surgimento de URF e CRF. Embora, no Brasil o subtipo C não seja a forma mais prevalente, três genótipos puros (B, C e F1) circulam na população que vive com HIV/AIDS (ARRUDA et al., 2018). A verificação dos pontos de recombinação entre B e C nas sequências, permitiu constatar que em três genes virais (*pol*, *env* e *nef*) há uma maior frequência destes eventos. Baseado nos dados de frequência de inserções gênicas comuns ao subtipo B, este subtipo parece apresentar características moleculares que conferem vantagens adaptativas, que está sendo transmitidas entre as cepas BC do HIV-1.

A busca por assinaturas moleculares nas sequências genoma de comprimento quase total pertencentes ao subtipo C do Brasil e do mundo, resultou na constatação de que grande parte das amostras deste subtipo não apresentam as características genéticas já descritas na literatura que estão associadas a uma maior condição patogênica associada ao subtipo C ou que favoreça uma maior capacidade de transmissão por esta cepa (ARENZANA-SEISDEDOS et al., 2006; FREED et al., 2007; SATISCHANDRA et al., 2000; SCRIBA et al., 2002; PIERSON & DOMS, 2003; WADIA et al., 2001). Embora tenham sido encontrados apenas 26 genomas no Brasil, a presença de assinaturas moleculares associadas a uma infecção menos agressiva foi observada e de acordo com as pesquisas de Graf e colaboradores (2021), Silveira e colaboradores (2012) e Souto e colaboradores (2021), contraditoriamente, estas características podem contribuir com uma maior chance de transmissão dessa forma viral, o que justifica um contínuo e discreto aumento do subtipo C nos estudos de epidemiologia molecular (GRAF et al., 2021).

## 6 CONCLUSÕES

- O subsubtipo F1 que circula no Brasil é descendente da linhagem I, enquanto as amostras da Romênia pertencem à linhagem II, onde estão alocadas a maioria das amostras africanas.
- As novas sequências F1, também agrupam na linhagem I. A análise de evolução populacional revelou uma diminuição desta expansão após os anos 2000. Isso evidencia ainda a contínua evolução deste subsubtipo, mesmo que no recombinante BF1, uma vez que as novas amostras BF1, quando analisadas apenas a região F1, agrupam próximas às demais brasileiras e todas no clado I.
- Os novos fragmentos e genoma virais do subtipo C obtidos através de coleta de amostras de pacientes acompanhados no Hospital Universitário Professor Edgard Santos, revelaram ser diretamente vinculado à região Sul. Agruparam-se filogeneticamente no clado principal brasileiro, fornecendo evidências que esse subtipo encontrado em território nordestino e baiano descende desta linhagem principal.
- As análises filodinâmicas revelaram múltiplas introduções do subtipo C no Brasil, com ancestralidade relacionada às formas de países do sul e leste do continente africano. Embora esse dado já tenha sido tratado em outro trabalho a partir da região Sudeste (DELATORRE et al., 2013), trazemos aqui uma nova evidência de chegada desta variante através da região Norte.
- Baseado nestas análises, foi possível estimar que o ancestral comum do subtipo C está presente em indivíduos na região Nordeste desde o meio da década de 80, através de múltiplos eventos de transição a partir das regiões Centro-oeste, Sudeste e Sul. Os dados sugerem movimentos desta cepa desde 1985 a 1998, podendo ainda ser verificados eventos de transmissão intraregionais, datados após o ano de 1985 (1986 e 1998).
- De acordo com as análises filogeográficas a forma viral do subtipo C que circula na região Nordeste é relacionada a cepa encontrada nas regiões ao sul do Brasil. De fato a literatura indica que este vírus alcançou território brasileiro através da região Sul, no estado do Paraná, constituindo, desta forma, a principal linhagem do subtipo C no Brasil.

- As análises moleculares destas amostras revelam assinaturas distintas e compartilhadas entre amostra da Bahia, quando comparada as sequências de outras localidades, sugerindo que, embora ainda em baixa prevalência, o subtipo C pode estar estabelecido nesta população.
- A amostra recombinante BC de nossa coorte revelou um novo padrão não antes descrito no estado da Bahia a nível do gene *pol*, região da integrase. Inserções do subtipo B em C nesta posição foram descritas ao sul e sudeste do país.
- As análises filogenéticas agrupam a sequência BC de nosso estudo com padrões de recombinantes BC de outros estados brasileiros (Rio Grande do Sul e Rio de Janeiro), sugerindo fortemente o desenvolvimento de uma nova CRF BC no país.
- Foram encontrados 225 genomas de comprimento quase total pertencentes ao recombinante BC na base de dados do Los Alamos. Dois países detêm o maior número de amostras e número de CRFs descritas: China e Brasil.
- A caracterização genômica revelou que há pontos preferenciais onde recombinação ocorre com mais frequência: *pol*, *env* e *nef*. A manutenção do subtipo B em C em posições específicas nestes genes parecem estar condicionado a vantagens moleculares, de acordo com os dados da literatura.
- As análises filogenéticas, deste trabalho, com as amostras F1 e BC revelam nós suportados relacionando sequências brasileiras e européias. Possivelmente há uma cadeia de transmissão envolvendo a América do Sul (Brasil) e Europa (Espanha e Itália). Eventos de transmissão do subsubtipo F1 do Brasil para Espanha foram observados (GUIMARÃES et al., 2009) e em concordância com Cañada et al., 2021 e Simonetti et al., 2014, amostras recombinantes BC brasileiras, espanholas (CRF108\_BC) e italianas (CRF60\_BC) agrupam todas juntas com as amostras brasileiras e sul-americanas (Argentina e Uruguai).
- As análises de assinaturas de aminoácidos utilizando 960 sequências nucleotídicas virais revelaram que a maioria das cepas C apresentam características moleculares associadas às que tem menor capacidade replicativa, embora essa propriedade biológica possa estar associada com maiores chances de transmissão.
- Avaliando as 26 sequências de genomas quase completos identificadas no Brasil, as características moleculares levantadas em nossa busca sugerem o mesmo comportamento de vírus encontrados em outras localidades do mundo, justificando as

baixas taxas de prevalência no país mas evidenciando o discreto aumento ao longo dos anos.

## REFERÊNCIAS

- ABBAS, A. K., LICHTMAN, A. H., PILLAI, S. *Imunologia celular e molecular*. 8.ed. - Rio de Janeiro: Elsevier, 2015.
- ALCALDE, R., GUIMARÃES, M. L., DUARTE, A. J. S., CASSEB, J. Clinical, epidemiological and molecular features of the HIV-1 subtype C and recombinant forms that are circulating in the city of São Paulo, Brazil. *Virology Journal*, 9:156, 2012.
- ALCÂNTARA, K. C., REIS, M. N. G., CARDOSO, L. P. V., BELLO, B., STEFANI, M. M. A. Increasing Heterosexual Transmission of HIV-1 Subtype C in Inland Central Western Brazil. *Journal of Medical Virology* 85:396–404, 2013.
- AMARAL, A. G., OLIVEIRA, I. B., CARNEIRO, D. C., ALCANTARA, L. C., & MONTEIRO-CUNHA, J. P. An overview of the molecular and epidemiological features of HIV-1 infection in two major cities of Bahia state, Brazil. *Memórias Do Instituto Oswaldo Cruz* 2017, 112(6), 411–418. doi:10.1590/0074-02760160458
- ANISIMOVA M, GASCUELA O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol* 2006; 55 (4): 539–552.
- ARRUDA, M. B., BOULLOSA, L. T., CARDOSO, C. C., DA COSTA, C. M., BRITES, C., DE LIMA, S. T. KAMINSKI, H. T., ALEIXO, A. W., ESPOSITO, O. P., CAVALCANTI, A. M., RIEDEL, M., COUTO-FERNANDEZ, J. C., FERREIRA, S. B., DE OLIVEIRA, I. C., PORTAL, L. E., WOLF, H. H., FERNANDES, S. B., PARDINI, M. I M. C., FEITEIRO, M. V., TOLENTINO, F. M., DIAZ, R. S., LOPES, G. I., FRANCISCO, R. B., VÉRAS, N. M., PIRES, A. F., FRANCHINI, M., MESQUITA, F., TANURI, A., HIV-BResNet. Brazilian network for HIV Drug Resistance Surveillance (HIV-BresNet): a survey of treatment-naïve individuals. *Journal of the International AIDS Society*, 21(3), e25032. 2018
- AULICINO, P. M., BELLO, G., ROCCO, C., ROMERO, H., MANGANO, A., MORGADO, M. G., SEN; L. Description of the First Full-Length HIV Type 1 Subtype F1 Strain in Argentina: Implications for the Origin and Dispersion of This Subtype in South America. *Aids Research and Human Retroviruses* Volume 23, Number 10, pp. 1176–1182, 2007.
- ARRUDA, E., SIMÕES, L., SUCUPIRA, M. C., MEDEIROS, M., ARRUDA, E., DIAZ, R. S., LIMA, A. Short communication: Intermediate prevalence of HIV type 1 primary antiretroviral resistance in Ceará state, Northeast Brazil. *AIDS Research and Human Retroviruses*, Volume 27, Number 2, 2011.
- AYEHUNIE S, JOHANSSON B, SONNERBORG A, SALMINEN M, ZEWDIE DW, FEHNIGERT, BRITTON S, STRANNEGARD O. New subtype of HIV-1 in Ethiopia. *Lancet*, 1990.
- BAENINGER, 2000. Africanos no Brasil: dubiedade e estereótipos. Migrações internacionais e identidade, Brasil. <http://www.comciencia.br/reportagens/migracoes/migr11.htm>.
- BALL, S. C., ABRAHA, A., COLLINS, K. R., MAROZSAN, A. J., BAIRD, H., QUINONES-MATEU, M. E., PENN-NICHOLSON, A., MURRAY, M., RICHARD, N., LOBRITZ,

M., ZIMMERMAN P. P., KAWAMURA, T., BLAUVELT, A., ARTS, E. J. Comparing the Ex Vivo Fitness of CCR5-Tropic Human Immunodeficiency Virus Type 1 Isolates of Subtypes B and C. *Journal of Virology*, 77(2), 2003.

BANDEA, C.I., RAMOS, A., PIENIAZEK, D., PASCU, R., TANURI, A., SCHOCHETMAN, G., RAYFIELD, M. A. 1995. Epidemiologic and evolutionary relationships between Romanian and Brazilian HIV-subtype F strains. *Emerg. Infect. Dis.* 1 (July–September (3)), 91–93.

BARRE-SINOUSSE, F., CHERMANN, J. C., REY, F., NUGEYRE, M. T., CHAMARERT, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, v. 220, n. 4599, p. 868–871, 20 maio 1983.

BELLO, G., PASSAES, C. P. B., GUIMARÃES, M. L., LORETE, R. S., ALMEIDA, S. E. M., MEDEIROS, R. M., ALENCASTRO, P. R., MORGADO, M. G. Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS*, Vol 22, No 15, 2008.

BELLO, G., WALTER, A., EYER-SILVA, J., COUTO-FERNANDEZ, J.C., GUIMARÃES, M.L., CHEQUER-FERNANDEZ, L.S., TEIXEIRA, S.L.M., MORGADO, M.G. Demographic history of HIV-1 subtypes B and F in Brazil. *Infection, Genetics and Evolution* 7 (2007) 263–270. <https://doi.org/10.1016/j.meegid.2006.11.002>

BELLO, G., ZANOTTO, P. M. DE A., IAMARINO, A., GRAF, T., PINTO, A. R., COUTO-FERNANDEZ, J. C., MORGADO, M. G. Phylogeographic Analysis of HIV-1 Subtype C Dissemination in Southern Brazil. *Plos One*. Issue. March 19, 2012.

BHANGE, D., PRASAD, N., SINGH, S., KUMAR, H., MAURYA, S. P., GOPALAN, B. P., NADIG, S. CHATURBHUIJ, D., JAYASEELAN, B., DINESHA, T. R., AHAMED, S. F., SINGH, N., BRAHMAIAH, A., MEHTA, K., GOHIL, Y., BALAKRISHNAN, P., DAS, B. K., DIAS, M., GANGAKHEDKAR R., MEHENDALE, S., PARANJAPE, R. S., SARAVANAN, S., SHET, A., SOLOMON, S. S., THAKAR, M. ., RANG, U. The Evolution of Regulatory Elements in the Emerging Promoter Variant Strains of HIV-1 Subtype C. *Frontiers in Microbiology*. 12:779472.

BOULLOSA, J., BACHU, M., BILA, M., RANGA, U., SÜFFERT, T., SASAZAWA, T., TANURI, A. Genetic Diversity in HIV-1 Subtype C LTR from Brazil and Mozambique Generates New Transcription Factor-Binding Sites. *Viruses*, 6,2495-2504, 2014.

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2017. Janeiro a junho 2017; Ano V, nº 01. Disponível em: <http://www.aids.gov.br/pt-br/pub/2017/boletim-epidemiologico-de-aids-2017>.

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2018. De julho 2017 a junho 2018; Ano V, volume 49, nº 53. Disponível em: <http://www.aids.gov.br/pt-br/pub/2018/boletim-epidemiologico-hivaid-2018>

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2019. 2019; nº especial. Disponível em: <http://www.aids.gov.br/pt-br/pub/2019/boletim-epidemiologico-de-hivaid-2019>

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2020. 2020; n° especial. Disponível em: <http://www.aids.gov.br/pt-br/pub/2020/boletim-epidemiologico-hivaid-2020>

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2021. 2021; n° especial. Disponível em: <http://www.aids.gov.br/pt-br/pub/2021/boletim-epidemiologico-hivaid-2021>

BRASIL. Ministério da Saúde. Boletim Epidemiológico de AIDS 2022. 2022; n° especial. Disponível em: <http://www.aids.gov.br/pt-br/pub/2022/boletim-epidemiologico-hivaid-2022>

BRÍGIDO, L. F., NUNES, C. C., OLIVEIRA, C. M., KNOLL, R. K., FERREIRA, J. L., FREITAS C. A., ALVES, M. A., DIAS, C., RODRIGUES, R., Research Capacity Program. HIV type 1 subtype C and CRF01\_AG recombinants prevail at the cities with the highest aids prevalence rate in Brazil. *AIDS Res Hum Retroviruses*. Dec; 23 (12): 1579-86, 2008.

CAÑADA, J.E., DELGADO, E., GIL, H., SÁNCHEZ, M., BENITO, S., GARCÍA-BODAS, E., GÓMEZ-GONZÁLEZ, C., CANUT-BLASCO, A., PORTU-ZAPIRAIN, J., DE ADANA, E.S., DE LA PENÁ, M., IBARRA, S., CILLA, G., IRIBARREN, J.A., MARTÍNEZ-SAPIÑA, A., THOMSON, M.M. Identification of a New HIV-1 BC Intersubtype Circulating Recombinant Form (CRF108\_BC) in Spain. *Viruses*. 13 (1): 93. 2021.

CARVALHO, B. C., CARDOSO, L. P. V., DAMASCENO, S., STEFANI, M. M. A. Moderate prevalence of transmitted drug resistance and interiorization of HIV Type 1 Subtype C in the Inland North State of Tocantins, Brazil. *AIDS Research and Human Retroviruses* Volume 27, Number 10, 2011.

CHERESHNEV, V.A., BOCHAROV, G., BAZHAN, S., BACHMETYEV, B., GAINOVA, I., LIKHOSHVAI, V., ARGILAGUET, J.M., MARTINEZ, J.P., RUMP, J.A., MOTHE, B., BRANDER, C., MEYERHANS, A. Pathogenesis and Treatment of HIV Infection: The Cellular, the Immune System and the Neuroendocrine Systems Perspective. *International Reviews of Immunology*, Early Online:1–25, 2013. <https://doi.org/10.3109/08830185.2013.779375>

CUNHA, L. K. H., KASHIMA, S., AMARANTE, M. F. C., HADDAD, R., RODRIGUES, E. S., SILVA, K. L. T., LIMA, T. A., CASTRO, D. B., BRITO, F. C., ALMEIDA, E. G., COVAS, D. T., MALHEIRO, A. Distribution of Human Immunodeficiency Virus type 1 subtypes in the State of Amazonas, Brazil, and subtype C identification. *Braz J Med Biol Res*, Volume 45 (2) 104-112, February 2012.

CLAPHAM, P. R., MCKNIGHT, A. Cell surface receptors, virus entry and tropism of primate lentiviruses. *Journal of General Virology*, 83, 1809– 1829, 2002.

CHIN, B. S., Molecular Epidemiology of Human Immunodeficiency Virus. *Infect Chemother* 49 :1-9, 2017.

DALAI, S.C., DE OLIVEIRA, T., HARKINS, G.W., KASSAYE, S.G., LINT, J., MANASA, J., JOHNSTON, E., KATZENSTEIN, D. Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. *AIDS* 2009; 23 (18):2523–32.

DA COSTA, C. M., DE OLIVEIRA, C. M. C, DE MELO, Y. F. C., DELATORRE, E., BELLO, G., COUTO-FERNANDEZ, J. C. High HIV-1 Genetic diversity in patients from Northern



Brazil. *AIDS Research and Human Retroviruses*, v.32, n.9, p.918-922, Sept. 2016.

DE OLIVEIRA, T., PILLAY, D., GIFFORD, R., for the UK Collaborative Group on HIV Drug Resistance (2010) The HIV-1 subtype C epidemic in south americais linked to the United Kingdom. *PLoS ONE* 5(2): e9311, February 19, 2010.

DELATORRE, E., COUTO-FERNANDEZ, J. C., BELLO, G. HIV-1 Genetic Diversity in Northeastern Brazil: high prevalence of non-B subtypes. *AIDS Research and Human Retroviruses*, 2017.

DELATORRE, E., COUTO-FERNANDEZ, J. C., GUIMARÃES, M. L., CARDOSO-VAZ, L. P., ALCANTARA, K. C. DE, STEFANI, M. M., ROMERO, H., FREIRE, C. C., IAMARINO, A., ZANOTTO, P. M. DE A., MORGADO, M. G., BELLO, G. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. *PLoS One*. 8 (9): e74072, February 2013.

DIMONTE, S., BABAKI-MINA, M., MERCURIO, F., DI PINTO, D., CECCHERINI-SILBERSTEIN, F., SVICHER, V., PERNO, C-F. Selected amino acid changes in HIV-1 subtype C go41 are associated with specific gp120V3 signatures in the regulation of co-receptor usage. *VirusResearch* 168, 73-83. 2012.

DRUMMOND, A. J., SUCHARD, M. A., XIE, D. & RAMBAUT, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969– 1973, 2012.

DRUMMOND, A.J., RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214. 2007.

DUMITRESCU, O., KALISH, M.L., KLIKS, S.C., BANDEA, C.I., LEVY, J.A. Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J Infect Dis.* 1994; 169(2): 281-8.

ESTATÍSTICAS UNAIDS, 2021.

Disponível em: [www.unaids.org.br/estatisticas/](http://www.unaids.org.br/estatisticas/) Acesso em: 13 de janeiro de 2022.

FABENI, L., BERNO, G., FOKAM, J., BERTOLI, A., ALTERI, C., GORI, C., FORBICI, F., TAKOU, D., VERGORI, A., ZACCARELLI, M., MAFFONGELLI, G., BORGHI, V., LATINI, A., PENNICA, A., MARIA, C., MONTELLA, M. F., MUSSINI, C., ANDREONI, M., ANTINORI, A., PERNO, C. F., SANTORO, M. M.; Comparative evaluation of subtyping tools for the surveillance of newly emerging HIV-1 strains. *J. Clin. Microbiology*. July, 2017.

FARIA, N.R., RAMBAUT, A., SUCHARD, M.A., BAELE, G., BEDFORD, T., WARD, M.J., ANDREW, J. T, SOUSA, J.D., ARINAMINPATHY, N., PÉPIN, J., POSADA, D., PEETERS, M., PYBUS, O.G., LEMEY, P. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014; 346 (6205) :56–61.

FARIA, N.R., VIDAL, N., LOURENCO, J., RAGHWANI, J., SIGALOFF, K. C. E., TATEM, A.J., VAN DE VIJVER, D. A. M., PINEDA-PEÑA, A. C., ROSE, R., WALLIS, C. L., AHUKA-MUDEKE, S., MUYEMBE-TAMFUM, C. J., MUWONGA, J., SUCHARD, M. A., DE WIT, T. R., HAMERS, R. L., NDEMBI, N., BAELE, G., PEETERS, M., PYBUS, O. G., LEMEY, P., DELLICOUR, S. al. Distinct rates and patterns of spread of the major HIV-1

subtypes in Central and East Africa. *PLoS Pathogens* 2019;15(12):e1007976.

FEIBERG, M. B., JARRET, R. F., ALDOVINI, A., GALLO, R. C., WONG-STAAAL, F. HTLV-III expression and production involve complex regulation at levels of splicing and translation of viral RNA. *Cell*, Vol. 46, 807-817, September 12, 1986.

FERRARO, G. A., MONTEIRO-CUNHA, J. P., FERNANDES, F. M. C., MOTA-MIRANDA, A. C. M., BRITES, C., ALCÂNTARA, L. C. J., CASTRO, B. G., MORGADO, M. G. Molecular characterization of Long Terminal Repeat sequences from Brazilian Human Immunodeficiency virus type 1 isolates. *AIDS Research and Human Retroviruses*. V. 29, N. 5, 2013.

FERREIRA, A. S., CARDOSO, L. P. V., STEFANI, M. M. A. Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. *Journal Medical of Virology*; 83, (8). 1301-1307. 2011

FRITSCH, H. M., ALMEIDA, S. E. M., PINTO, A. R., GRÄF, T. Spatiotemporal and demographic history of the HIV-1 circulating recombinant form CRF31\_BC in Brazil. *Infection, Genetics and Evolution*, 61, 113–118. 2018.

GAO, F., BAILES, E., ROBERTSON, D. L., CHEN, Y., RODENBURG, C. M., MICHAEL, S. F., CUMMINSK, L. B., ARTHUR, L. O., PEETERS, M., SHAW, G. M., SHARP, P. M., HAHN, B. H. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature*, v. 397. 4 February 1999.

GARTNER, M. J., ROCHE, M., CHURCHILL, M. J., GORRY, P. R., FLYNN, J. K., Understanding the mechanisms driving the spread of subtype C HIV-1. *eBiomedicine*. 53, 102682.

GAO, Q., FENG, Y., GAO, L., ZHANG, M., LIU, Y., XIAO, M., DONG, X., XIA, X. Genetic characteristics of a New HIV-1 subtype B/C intersubtype circulating recombinant form (CRF118\_BC) identified in Yunnan, China. *Journal of Infection*. 83 (2): 237-279. 2021.

GELDERBLOM, H. R., OZEL, M., HAUSMANN, E. H. S., WINKEL, T., PAUL, G., KOCH, M. A. Fine structure of Human Immunodeficiency Virus (HIV), immunolocalization of structural proteins and virus-cell relation. *Micron and Microscopica*. Vol. 19, No. 1, pp. 41—60, 1988.

GRÄF, T., PINTO, A. R. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 435, 170–178, 2013.

GRÄF, T., VRANCKEN, B., JUNQUEIRA, D. M., MEDEIROS, R. M. DE, SUCHARD, M. A., LEMEY, P., ALMEIDA, S. E. DE M., PINTO, A. R. Contribution of Epidemiological Predictors in Unraveling the Phylogeographic History of HIV-1 Subtype C in Brazil. *Journal of Virology*. Volume 89 Number 24; December 2015.

GRÄF, T., BELLO, G., ANDRADE, P., ARANTES, I., PEREIRA, J. M., DA SILVA, A. B. P., VEIGA, R. V., MARIANI, D., BOULLOSA L, T., ARRUDA, M, B., FERNANDEZ, J, C, C., DENNIS, A, M., RASMUSSEN D, A., TANURI, A. (2021). HIV-1 molecular diversity in

Brazil unveiled by 10 years of sampling by the national genotyping network. *Scientific Reports*, 11(1). doi:10.1038/s41598-021-94542-5

GIOVANETTI, M., CICCOCCHI, M., PAROLIN, C., BORSETTI, A. *Molecular Epidemiology of HIV-1 in African Countries: A Comprehensive Overview*. *Pathogens*, 9, 1072. 2020.

GUIMARÃES, M.L., VICENTE, A.C.P., OTSUKI, K., SILVA, R.F.F.C., FRANCISCO M., MORGADO, M.G., BELLO, G. Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology*. 2009. <https://doi.org/10.1186/1742-4690-6-39>

HALL, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 1999.

HAN, X., AN, M., ZHAO, B., DUAN, S., YANG, S., XU, J., ZHANG, M., MCGOOGAN, J.M., TAKEBE, Y., SHANG, H. High Prevalence of HIV-1 Intersubtype B9/C Recombinants among Injecting Drug Users in Dehong, China. *PLoS One*. 31: 8 (5) e65337.2013.

HARRICH, D., ULLICH, C, L.GARCIA-MARTÍNEZ, F., GAYNOR, R. B. Tat is required for efficient HIV-1 reverse transcription. *The EMBO Journal* Vol.16 No.6 pp.1224–1235, 1997.

HENG, X., TIAN, C., CHOI, K.H., ZHANG, J., CHENG, H., YANG, X., LI, D., LIN, J., QU, S., SUN, X. Injecting drug use and HIV infection in southwest China. *AIDS*. 8: 1141–1147. 1994.

HEMELAAR, J. The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*, 18(3), 182–192. 2012.

HUANG, Y., NIU, B., GAO, Y., FU, L., LI, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, n5, pp. 680-682. 2010

HUNT, G. M., JOHNSON, D., TIEMESSE, C. T. Characterization of the long terminal repeat regions of South African human immunodeficiency virus type 1 isolates. *Virus Genes* 2001.

JACKSON, J. B., BALFOUR-JR, H. H. *Practical Diagnostic Testing for Human Immunodeficiency Virus*. *Clinical Microbiology Reviews*, p. 124-138, January 1988.

IOACHIM, H. L. Immunopathogenesis of Human Immunodeficiency Virus infection. *Cancer Research* 1990; 50: 5612-5617

JONES, L. R., DILERNIA, D. A., MANRIQUE, J. M., MORETTI, F., SALOMÓN, H., GOMEZ-CARRILLO, M. In-Depth Analysis of the Origins of HIV Type 1 Subtype C in South America. *Aids Research and Human Retroviruses* Volume 25, Number 10, 2009.

KALU, A., W., TELELE, N., F., GEBRESELASIE, S., FEKADE, D., ABDURAHMAN, S., MARRONE, G., SONNERBORG, A. Monophyletic HIV-1C epidemic in Ethiopia is dominated by CCR5-tropic viruses-an analysis of a prospective country- wide cohort. *BMC Infectious Diseases*, 17:37. 2017.

KALYAANAMOORTHY, S., MINH, B., WONG, T., HAESELER, A., JERMIIN, L. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*. 14 (6): 587–589. 2017.

KATNER, H. P., PANKEY, G. A. Evidence for a euro-american origin of human immunodeficiency virus (HIV). *Journal of the National Medical Association*, vol. 79, N<sup>o</sup>. 10, 1987.

KATOH K., KUMA K., TOH H., MIYATA T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33, 511–8, 2005.

KEARSE, M., MOIR, R., WILSON, M., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P., DRUMMOND, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics Applications Note*. Vol. 28, pages 1647–1649. no. 12 2012.

KIGUOYA, M.W., MANN, J.K., CHOPERA, D., GOUNDER, K., LEE, G.Q., HUNT, P.W., MARTIN, J. N., BALL, B., KIMANI, J., BRUMME, Z. L., BROCKMAN, M.A., NDUNG'U, T. Subtype Specific Differences in Gag-Protease-Driven Replication Capacity Are Consistent with Intersubtype Differences in HIV-1 Disease Progression. *Journal of Virology*; 91(13) e00253-17. 2017.

LI, J., GAO, Q., ZHANG, M., LIU, J., JIA, Y., FENG, Y., XIA, X., DONG, X. A newly emerging HIV-1 circulating recombinant form (CRF110\_BC) comprising subtype B and C among intravenous drug users in Yunnan, China. *Journal of Infection*. 82 (3): e8–e10. 2021.

LIMA, K., LEAL, E.S., CAVALCANTI, A.M.S., SAULISTIANO, D.M., MEDEIROS, L.B., SILVA, S.P., LACERDA, H.R. Increase in Human Immunodeficiency Virus 1 diversity and detection of various subtypes and recombinants in north-eastern Brazil. *J Med Microbiol*. 2017; 66(4): 526-535. <https://doi.org/10.1099/jmm.0.000447>

LIMA, K., LEAL, E.S., CAVALCANTI, A.M.S., SALUSTIANO, D.M., MEDEIROS, L.B., SILVA, S.P., LACERDA, H.R. Epidemiological, Clinical and Antiretroviral Susceptibility Characterization of Human Immunodeficiency Virus Subtypes B and Non-B in Pernambuco, Northeast Brazil. *PLOS ONE*. May 24, 2016. <https://doi.org/10.1371/journal.pone.0155854>

LINNER, K. J., HALL, C. D., ROBERTSON, K. R. Effects of antiretroviral therapy on cognitive impairment. *AIDS/HIV Rep*. 5:64. 2008.

LOUWAGIE, J., DELWART, E.L., MULLINS, J.I., MCCUTCHAN, F.E., EDDY, G., BURKE, D.S. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res Hum Retroviruses*. 1994; 10: 561–567.

MACHADO, L. F. A., COSTA, I. B., FOLHA, M. N., DA LUZ, A. L. B., VALLINOTO, C. R., ISHAK, R., ISHAK, M. O. G. Lower genetic variability of HIV-1 and antiretroviral drug resistance in pregnant women from the state of Pará, Brazil. *BMC Infectious Diseases*, 17:270, 2017.

MARTIN, D. P., VARSANI, A., ROUMAGNAC, P., BOTHA, G., SURESH, M., SCHWAB,

KELZ, Z., KUMAR, V., MURREL, B. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*; 7. 2021

MELO, F. L. Caracterização biológica e molecular de recombinantes naturais de HIV-1. 2011. Tese (Doutorado em Microbiologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2011.

MEHTA, S.J., WERTHEIM, J.O., DELPORT, W., ENE, L., TARDEI, G., DUICULESCU, D., KOSAKOVSKY, S.L., SMITH, DM. Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. *Infects, Genetics and Evolution*; 11(5):975-9. 2011. <https://doi.org/10.1016/j.meegid.2011.03.009>

METHA, G., SHARMA, A., ARORA, S. K. Acquisition of Additional Nuclear Factor Kappa B Binding Sites in Long Terminal Repeat of Genetically Evolving HIV-1 Subtype C Viral Species in Host with Comorbidities. *AIDS Res Hum Retroviruses*. 2021 May;37(5):380-384.

MISHRA, M., VARGHESE, R. K., VERMA, A., DAS, S., AGUIAR, R. S., TANURI, A., MAHADEVAN, A., SHANKAR, S. K., SATISCHANDRA, P., RANGA, U. Genetic diversity and proviral DNA load in different neural compartments of HIV-1 subtype C infection. *Journal of Neurovirology*. 2015.

MONTEIRO, J. P., ALCANTARA, L. C. J., OLIVEIRA, T. DE, OLIVEIRA, A. M., MELO M. A. G., BRITES, C., GALVÃO-CASTRO, B. Genetic Variability of Human Immunodeficiency Virus-1 in Bahia State, Northeast, Brazil: High Diversity of HIV Genotypes. *Journal of Medical Virology* 81:391–399. 2009.

MONTEIRO-CUNHA, J. P., ARAUJO, A. F., SANTOS, E., GALVAO-CASTRO, B., ALCANTARA, L. C. J. Lack of High-Level Resistance Mutations in HIV Type 1 BF Recombinant Strains Circulating in Northeast Brazil. *AIDS Research and Human Retroviruses*. Volume 27, Number 6. 2011.

MONTEIRO-CUNHA, J. P., MOTA DE SOUZA, J. S., DIAZ, R. S., REGO, F. F. DE A., ANDRADE, E., SILVA, M. O., GIOVANETTI, M., OLIVEIRA, T. DE, OLIVEIRA, A., BRITES, R. C., ALCÂNTARA, L. C. HIV-1 Molecular Characterization and Transmitted Drug Resistance Prevalence among Treatment-Naïve Individuals. *Journal of AIDS and HIV Infections*. Volume 3 | Issue 1. 2017.

MOURA, M. E. S., REIS, M. N. G., LIMA, Y. A. R., CARDOSO, L. P. V., STEFANI, M. M. A. Low Rate of Transmitted Drug Resistance May Indicate Low Access to Antiretroviral Treatment in Maranhão State, Northeast Brazil. *AIDS Res Hum Retroviruses*; 31(2): 250–254. 2015.

MURRELL, B., MOOLA, S., MABONA, A., WEIGHILL, T., SHEWARD, D., KOSAKOVSKY, S.L., SCHEFFLER, L. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol*. 2013. May;30(5):1196-205. <https://doi.org/10.1093/molbev/mst030>.

NICHOLAS, K. B. GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNET NEWS*. 1997; 4:14.

NGUYEN, LT., SCHMIDT, H. A., HAESLER, A. V., MINH, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*; 32 (1): 268–274. 2015.

NOMAGUCHI, M., DOI, N., KOMA, T., ADACHI, A. HIV-1 mutates to adapt influxing environments. *Microbes and Infection* xx, 1 e 5, 2017.

OLIVEIRA, R. C., GRAF, T., REGO, F.F.A., SILVA, G. P. S. A., GIOVANETTI, M., MONTEIRO-CUNHA, J. P. Dynamic Dispersion of HIV-1 Subtype C Toward Brazilian Northeastern Region. *AIDS Research and Human Retroviruses* 2021; 37 (12): 913-921. <https://doi.org/10.1089/aid.2020.0308>

PASSAES, C.P.B., GUIMARÃES, M.L., BELLO, G., MORGADO, M.G. Near Full-Length Genome Characterization of HIV Type 1 Unique BC Recombinant Forms from Southern Brazil. *Aids Research and Human Retroviruses*. 2009. 25: 12.

PEREIRA, E. A., DA SILVA, L. P. HIV-1 Nef: Taking Control of Protein Trafficking. *Traffic*. 17; 9: 976-996. <https://doi.org/10.1111/tra.12412>.

PERELSON, A. S., NEUMANN, A. U., MARKOWITZ, M., LEONARD, J. M., HO, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 271(5255):1582–6. 1996.

PERNAS, B., GRANDAL, M., MENA, A., CASTRO-IGLESIAS, A., CAÑAZARES, A., WYLES, D.L., LÓPEZ-CALVO, S., PÉRTEGA, S., RODRÍGUEZ-OZORIO, I., PEDREIRA, J.D., POVEDA, E. High prevalence of subtype F in newly diagnosed HIV-1 persons in northwest Spain and evidence for impaired treatment response. *AIDS*; 2014 Jul 31;28(12):1837-40. <https://doi.org/10.1097/QAD.0000000000000326>.

PESSÔA, R., LOUREIRO, P., LOPES, M. E., CARNEIRO-PROIETTI, A. B. F., SABINO, E. C., BUSCH, M. P. SANABANI, S. S. Ultra-Deep Sequencing of HIV-1 near Full-Length and Partial Proviral Genomes Reveals High Genetic Diversity among Brazilian Blood Donors. *PLoS One*. 2016; 11(3): e0152499. <https://doi.org/10.1371/journal.pone.0152499>

RAMBAUT, A., LAM, T. T., FAGUNDES DE CARVALHO, L., PYBUS, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2, vew007 2016.

REIS, M. N. G., GUIMARÃES, M. L., BELLO, G., STEFANI, M. M. A. Identification of new HIV-1 circulating recombinant forms CRF81-cpx and CRF99-BF1 in central Western Brazil and of unique BF1 recombinant forms. *Frontiers in Microbiology*, v. 10, n. FEB, p. 1–13, 2019.

ROBERTSON, D. L., ANDERSON, J.P., BRADAC, J.A. HIV-1 nomenclature proposal. *Science* 2000; 288 (5463): 55-56. <https://doi.org/10.1126/science.288.5463.55d>

SALEMI, M., OLIVEIRA, T. DE, SOARES, M. A., PYBUS, O., DUMANS, A. T., VANDAMME, A. M., TANURI, A., CASSOL, S., FITCH, W. M. Different Epidemic Potentials of the HIV-1B and C Subtypes. *J. Mol. Evol.* 60:598–605, 2005.

SALMINEN, M. O., JOHANSSON, B., SONNERBORG, A., AYEHUNIE, S., GOTTE, D., LEINIKKI, P., BURKE, D. S., MCCUTCHAN, E. Full-length sequence of an Ethiopian Human Immunodeficiency Virus type-1 (HIV-1) isolate of genetic subtype C. *AIDS Research and Human Retroviruses*. Vol. 12, n<sup>o</sup>. 14, 1996.

SANABANI, S.S., PASTENA, E.R.S., COSTA, A.C., MARTINEZ, V.P., KLEINE-NETO, W., DE OLIVEIRA, A.C.S, SAUER, M.M., BASSICHETO, K.C., OLIVEIRA, S.M.S., TOMIYAMA, H.T.I., SABINO, E.C., KALLAS, E.G. Characterization of Partial and Near Full-Length Genomes of HIV-1 Strains Sampled from Recently Infected Individuals in São Paulo, Brazil. *Plos One*. 6: 10. e25869. 2011.

SANABANI, S., NETO, W.K., KALMAR, E.M.N., DIAZ R.S., JANINI, L.M., SABINO, E.C. Analysis of the near full-length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in São Paulo, Brazil. *Infection, Genetics and Evolution* 6 (2006) 368–377. <https://doi.org/10.1016/j.meegid.2006.01.003>

SANTORO, M. M., PERNO, C. F. HIV-1 Genetic Variability and Clinical Implications. Hindawi Publishing Corporation ISRN. Microbiology. 20 pages, 2013.

SANTOS, A.F., SOUSA, T.M., SOARES, E.A.J.M., SANABANI, S., MARTINEZ, A.M.B., SPRINZ, E., SILVEIRA, J., SABINO, E.C., TANURI, A., SOARES, M.A. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS*. 20; (16): 2011-2019. 2006

SANTOS, A.F., SCHRAGO, C.G., MARTINEZ, A.M.B., MENDOZA-SASSI, R., SILVEIRA, J., SOUSA, T.M., LENGROBER, R.B., SOARES, E. A J.M., SPRINZ, E., SOARES, M. Epidemiologic and evolutionary trends of HIV-1 CRF31\_BC-related strains in southern Brazil. *J. Acquir. Immune Defic. Syndr*. 45, 328–333. 2007.

SATISHCHANDRA, P., A. NALINI, M. GOURIE-DEVI, N. KHANNA, V. SANTOSH, V. RAVI, A. DESAI, A. CHANDRAMUKI, P. N. JAYAKUMAR, AND S. K. SHANKAR. Profile of neurologic disorders associated with HIV/AIDS from Bangalore, south India. *Indian J. Med. Res*. 111:14-2. 2000.

SCHMIDT, H.A., E. PETZOLD, M. VINGRON, AND A. VON HAESLER. Molecular Phylogenetics: Parallelized Parameter Estimation and Quartet Puzzling. *J. Parallel Distrib. Comput*. 63, 719-72730, 2003.

SCHULTZ, A, K.; ZANG, M.; BULLA, I.; LEITNER, T.; KOBBER, B.; MORGENSTERN, B.; STANKE, M. jpHMM: Improving the reability of recombination prediction in HIV-1. *Nucleic Acids Researc*. 1; 37. 2009.

SCRIBA, T. J., VILLIERS, T., TREURNICHT, F. K., MEGEDE, J. Z., BARNETT, S. W., ENGELBRECHT, S., RESENBURG, E. J. V. Characterization of the South African HIV type 1 Subtype C complete 5' Long Terminal Repeat nef and regulatory genes. *AIDS Research and Human Retroviruses*. V. 18. N.2. PP 149-159. 2002.

SHARP, P. M., BAILES, E., CHAUDHURI, R. R., RODENBURG, C. M., SANTIAGO, M, O., HAHN, B. H. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B Biol Sci*. Jun 29; 356(1410): 867–876. 2001.

SILVEIRA, J., SANTOS, A. F., MARTÍNEZ, A. M. B., GÓES, L. R., MENDOZA- SASSI, R., MUNIZ, C. P., TUPINAMBÁS, U., SOARES, M. A., GRECO, D. B. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *Journal of Clinical Virology*. N. 54, 36– 41, 2012.

SILVA, G. P. S. A., OLIVEIRA, R. C., SOUZA, J. S. M., GIOVANETTI, M., GUIMARÃES, M. L., BRITES, C., MONTEIRO-CUNHA, J. P. Tracing the relationship among HIV-1 sub-subtype F1 strains: a phylodynamic perspective. *Memórias do Instituto Oswaldo Cruz*. V. 117; Rio de Janeiro, 2022. [https://doi.org/ 10.1590/0074-02760220109](https://doi.org/10.1590/0074-02760220109)

SIMONETTI, F.R., LAI, A., MONNO, L., BINDA, F., BRINDICCI, G., PUNZI, G., BZZI, G., VIOLIN, M., GALLI, M. ZAZZI, M., ANGARANO, G., BOLATTA, C. Identification of a new HIV-1 BC circulating recombinant form (CRF60<sub>BC</sub>) in Italian young men having sex with men. *Infection, Genetics and Evolution*. 23: 176–181. 2014.

SOUTO, B., TRIUNFANTE, V., SANTOS-PEREIRA, A., MARTINS, J., ARAÚJO, P. M. M., OSÓRIO, N. S. Evolutionary dynamics of HIV-1 subtype C in Brazil. *Sci Rep.*; 11: 23060. 2021.

SOUSA, L.R.M., ELIAS, H.C., FERNANDEZ, N.M., GIR, E., REIS, R.K. Knowledge of PEP and PrEP among people living with HIV/aids in Brazil. *BMC Public Health*. 2021; 21: 64. Published online. Jan 7. 2021. doi: 10.1186/s12889-020-10135-3

SOUZA, J. S. M., JÚNIOR, J. J. S., BRITES, C. MONTEIRO-CUNHA, J. P. Molecular and geographic characterization of HIV-1 BF recombinants viruses. *Virus Research*. 2019; 270;197650. doi: 10.1016/j.virusres.2019.197650.

STEFANI, M.M.A, PEREIRA, G.A.S., LINS, J.A.B., ALCANTARA, K.C., SILVEIRA, A.A.V., MAYA, N.C., MUSSI, A.H. Molecular screening shows extensive HIV-1 genetic diversity in Central West Brazil. *Journal of Clinical Virology* 39 (2007) 205–209. <https://doi.org/10.1016/j.jcv.2007.04.012>

SU, L., GRAF, M., ZHANG, Y., VON BRIESEN, H., KOSTLER, J., MELZL, H., WOLF, H., SHAO, Y., WAGNER, R. Characterization of a Virtually Full-Length Human Immunodeficiency Virus Type 1 Genome of a Prevalent Intersubtype (C/B') Recombinant Strain in China. *Journal of Virology*. 74 (23): 11367–11376. 2000.

TRIFINOPOULOS, J., NGUYEN, L. T., HAESELER, A. V., MINH. B. Q. W-IQ- TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016.

VÉRAS, N. M. C., GRAY, R. R., BRÍGIDO, L. F. DE M., RODRIGUES, R., SALEMI, M. High-resolution phylogenetics and phylogeography of Human Immunodeficiency Virus type 1 subtype C epidemic in South America. *Journal General Virology*, 92. Pp. 1698-1708, 2011.

WADIA, R., PUJARI, S., KOTHARI, S., UDHAR, M., KULKARNI S., BHAGAT, S., NANIVADEKAR, A. Neurological manifestations of HIV disease. *Journal of Associations Physicians India*. 49: 343-348. 2001.



YAMAGUCHI, J., MCARTHUR, C., VALLARI, A., STHRESHLEY, L., CLOHERTY, G.A., BERG, M.G., RODGERS, M.A. Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *Journal of Acquired Immune Deficiency Syndromes*. 83: (3); 319-322. 2019.

ZHANG, M., LI, H., ZOU, Y., FU, Y., REN, Q., CHANG, W. Identification of Three HIV-1 Unique (B/C) Recombinant Forms in Shaanxi, China. *AIDS Research and Human Retroviruses*. 35 (9): 876-880. 2019.

## APÊNDICES

### APÊNDICE A - termo de consentimento livre-esclarecido



Instituto de Ciências da Saúde  
Universidade Federal da Bahia



#### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

O Sr. (a) está sendo convidado (a) como voluntário (a) a participar da pesquisa **“Identificação e Caracterização Molecular de Genomas Completos de Formas Recombinantes Circulantes do Vírus da Imunodeficiência Humana do tipo 1 (HIV-1) Circulantes na Bahia”**. Nesta pesquisa pretendemos identificar os subtipos do HIV-1 mais encontrados, investigar a existência de uma forma genética do vírus ainda não identificada no estado da Bahia, estimar de que local e quando esse agente viral foi trazido e identificar quais e o número de mutações que podem tornar o HIV resistente aos medicamentos”. Esta pesquisa está sendo realizada pois estudos anteriores indicaram que uma forma genética do vírus ainda não classificada está presente na Bahia. Para esta pesquisa adotaremos os seguintes procedimentos: será realizada a coleta de sangue por punção endovenosa asséptica em tubos monovetes contendo anticoagulante. A partir de sua amostra de sangue serão realizados os estudos necessários que identificarão as características do vírus e desta forma tentar esclarecer os objetivos do estudo acima citados. A sua participação no estudo será através do preenchimento de um questionário e o senhor(a) pode se negar a responder quaisquer questões, e da autorização do uso e do armazenamento do material sanguíneo excedente coletado para o estudo da carga viral e CD4. O conhecimento do subtipo específico com o qual está infectado provavelmente não trará qualquer utilidade para o seu tratamento, mas poderá ajudar na compreensão de como o vírus está se disseminando pela Bahia. Será pesquisada a presença de mutações virais que o tornam mais resistente aos medicamentos que usa ou poderá vir a usar para combatê-lo. Estas informações, sim, poderão ter grande valia para o seu tratamento. Os riscos pela participação neste estudo são mínimos, e se limitam a eventuais complicações no local da coleta de sangue, não sendo necessário coleta de amostra sanguínea maior do que a já prevista para os exames solicitados pelo meu médico (carga viral e CD4). Para evitar o aparecimento de hematoma após a punção venosa, recomenda-se: pressionar o local por no mínimo três minutos após a punção, evitar flexionar o braço, não massagear o local da coleta, não fazer esforço físico com o braço puncionado durante a primeira hora e evitar roupa, relógio ou pulseiras garroteando o braço puncionado. Em caso de aparecimento de hematoma, as seguintes medidas que devem ser tomadas: colocar compressas de gelo por



Instituto de Ciências da Saúde  
Universidade Federal da Bahia



15 minutos a cada hora nas primeiras seis horas. Em seguida podem ser colocadas compressas mornas para ajudar a eliminar o hematoma mais rapidamente. Em casos de formigamento e sensação de perda de força nas mãos, deve-se entrar em contato com o Laboratório. A coleta será realizada de forma asséptica, por profissional treinado em biossegurança, adotando equipamentos de proteção adequados (luvas, máscara e avental). Após a coleta, o material perfurocortante será descartado em recipiente rígido e resistente à perfurações sem reencapar. Para participar deste estudo o Sr (a) não terá nenhum custo, nem receberá qualquer vantagem financeira. Apesar disso, caso seja identificado e comprovado danos provenientes desta pesquisa, o Sr.(a) tem assegurado o direito a indenização a ser definida por meio de decisão judicial ou extrajudicial. Terá o esclarecimento sobre o estudo em qualquer aspecto que desejar e estará livre para participar ou recusar-se a participar. Poderá retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que é atendido pelo pesquisador, que tratará a sua identidade com padrões profissionais de sigilo.

Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. O (A) Sr (a) não será identificado em nenhuma publicação que possa resultar. Este termo de consentimento encontra-se impresso em duas vias originais, sendo que uma será arquivada pelo pesquisador responsável, no "**Instituto de Ciências da Saúde, Universidade Federal da Bahia**" e a outra será fornecida ao senhor. Os dados e instrumentos utilizados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos, e após esse tempo serão destruídos. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, atendendo a legislação brasileira (Resolução Nº 466/12 do Conselho Nacional de Saúde), utilizando as informações somente para os fins acadêmicos e científicos.

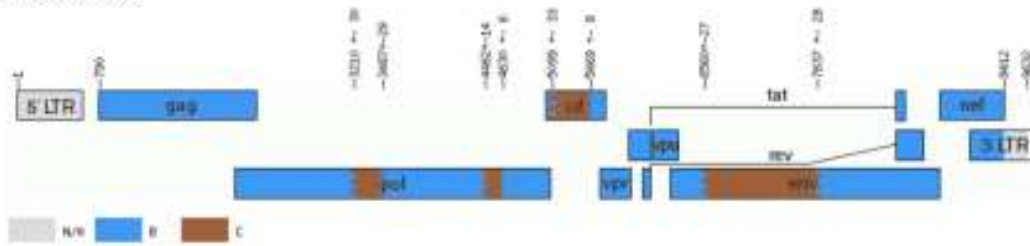
Eu, \_\_\_\_\_, portador do documento de Identidade \_\_\_\_\_ fui informado (a) dos objetivos da pesquisa "**Identificação e Caracterização Molecular de Genomas Completos de Formas Recombinantes circulantes do Vírus da Imunodeficiência Humana do tipo 1 (HIV-1) Circulantes na Bahia.**", de maneira clara e detalhada e esclareci minhas dúvidas. Sei que a

## APÊNDICE B – Questionário

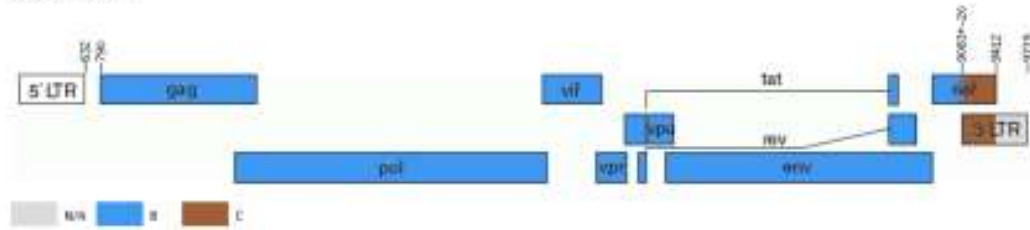
<b>Nome do paciente:</b>
<b>Médico:</b>
<b>Prontuário:</b>
<b>Idade:</b>
<b>Gênero:</b>
<b>Raça:</b>
<b>Data de nascimento:</b>
<b>Endereço:</b>
<b>Telefone:</b>
<b>Faixa salarial:</b>
<b>Escolaridade:</b>
<b>Data do diagnóstico:</b>
<b>Data da infecção:</b>
<b>Estado clínico:</b> ( ) Sintomático ( ) Assintomático
<b>Em tratamento (atual)?</b> ( ) Sim ( ) Não <b>Qual?</b>
<b>Tratamento anterior?</b> ( ) Sim ( ) Não <b>Qual?</b>
<b>Via de transmissão:</b> ( ) SH ( ) SM ( ) SB ( ) Transfusão ( ) Vertical ( ) Drogas injetáveis
<b>Orientação sexual:</b> ( ) Heterossexual ( ) Bissexual ( ) Homossexual ( ) Outros ( ) Não informado
<b>Usuário de drogas injetáveis?</b> ( ) Sim ( ) Não ( ) Não informado
<b>Utiliza outras medicações? Quais?</b>
<b>CD4 no diagnóstico:</b>
<b>CD4 atual:</b>
<b>Carga viral no diagnóstico:</b>
<b>Carga viral atual:</b>
<b>Outras infecções:</b> ( ) HTLV ( ) Tuberculose ( ) Sífilis ( ) Hepatite B ( ) Hepatite C ( ) Outros

**APÊNDICE C - Recombination patterns of HIV-1 BC samples classified as URFs used in frequency data of study: GENOMIC CHARACTERIZATION OF HIV-1 BC RECOMBINANT VIRUSES. Oliveira, R. C. and Monteiro-Cunha, J. P. (2023).**

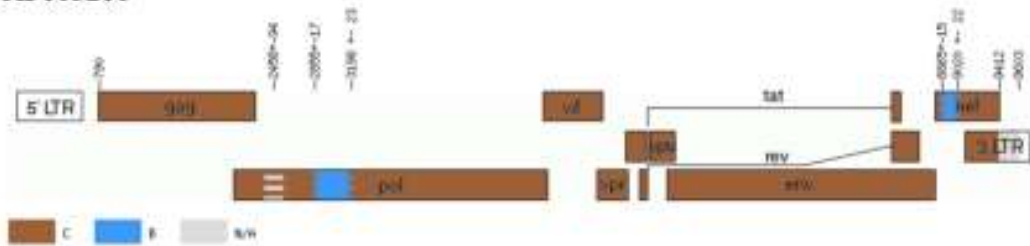
AB097869



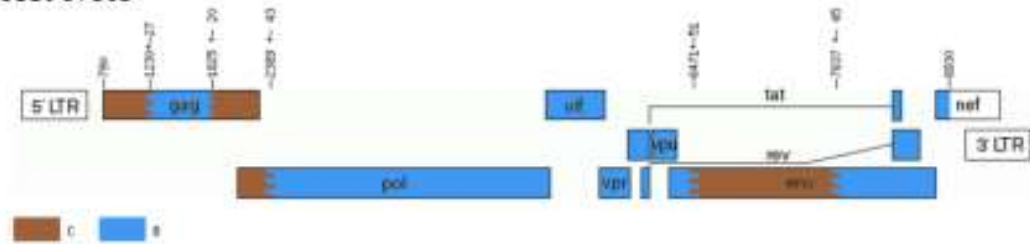
AB097873



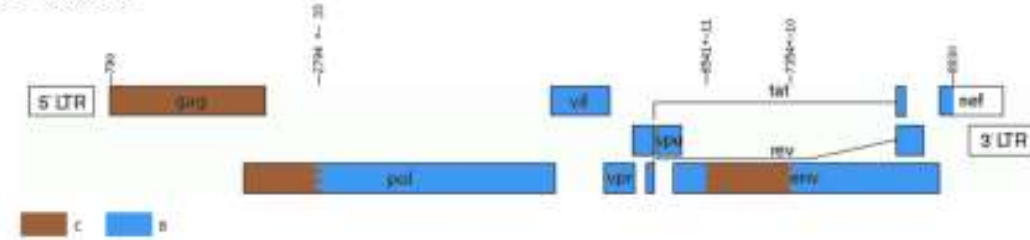
AP005206



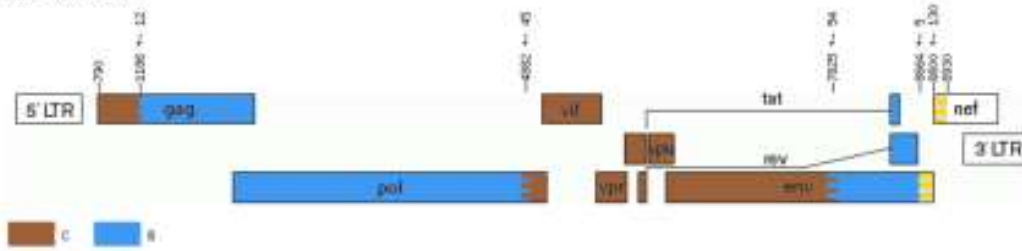
AY967803



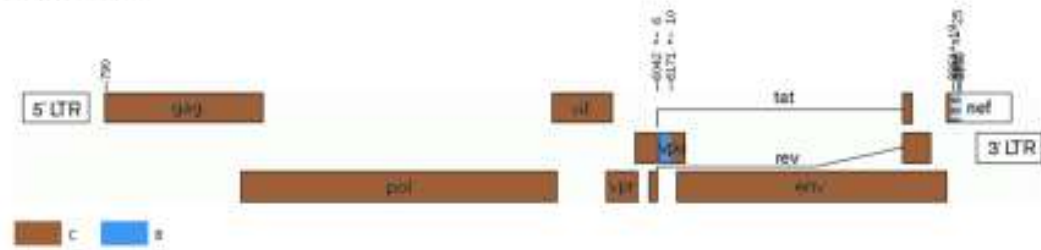
AY967804



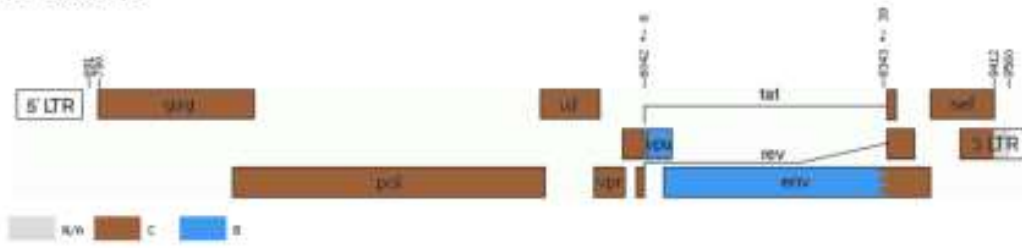
AY967805



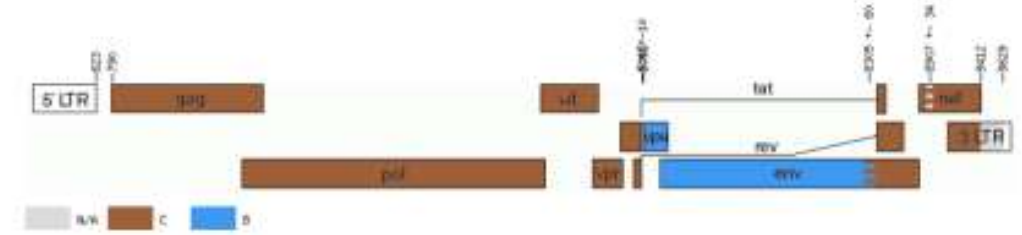
AY967807



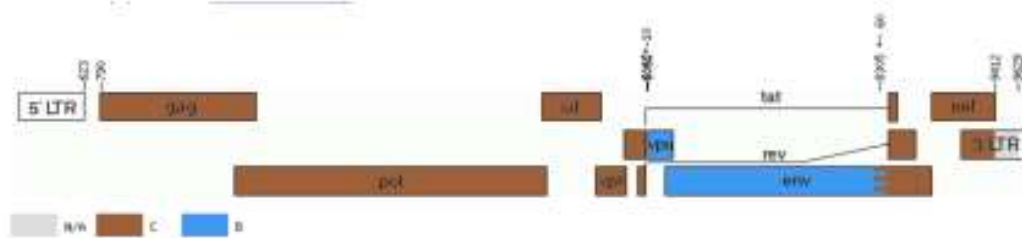
AY968312



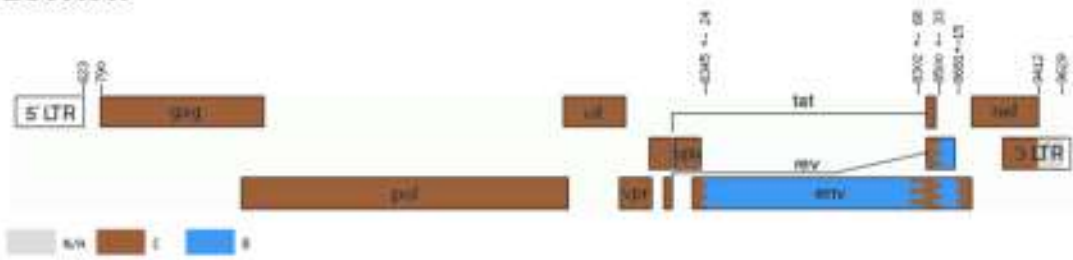
EU000506



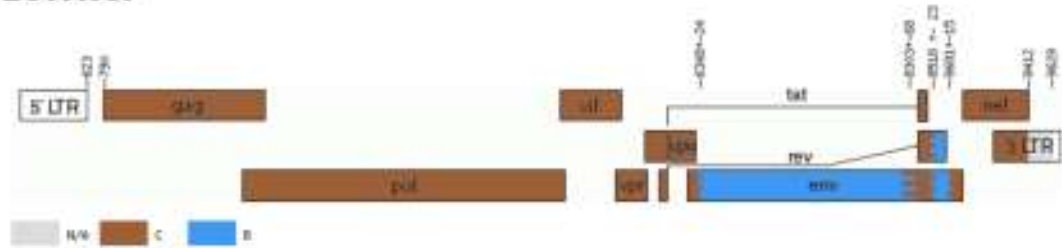
EU000508



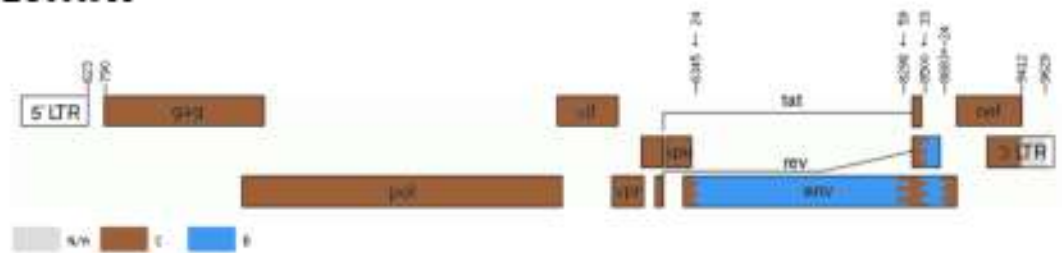
EU000509



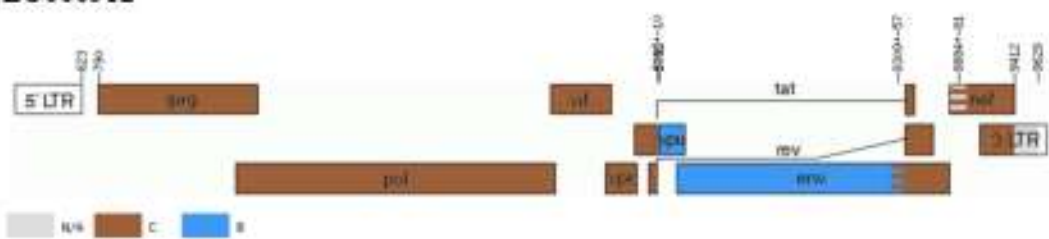
EU000510



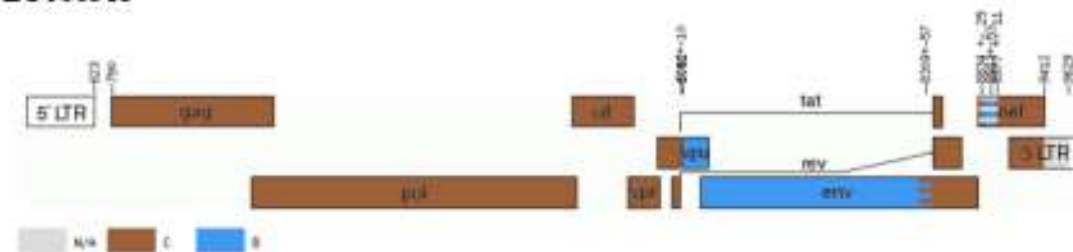
EU000511



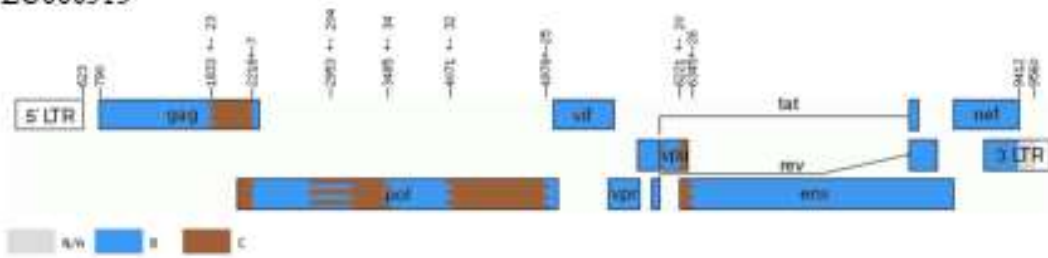
EU000512



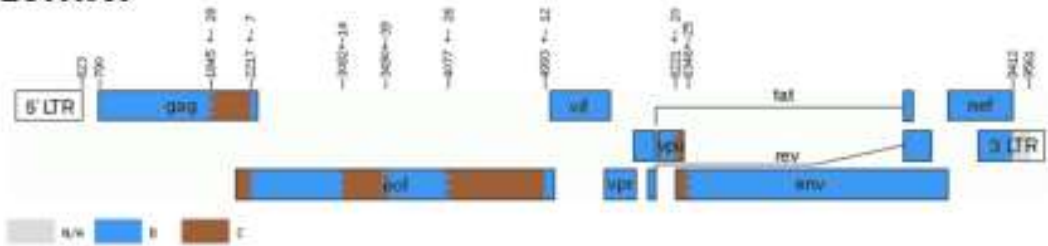
EU000513



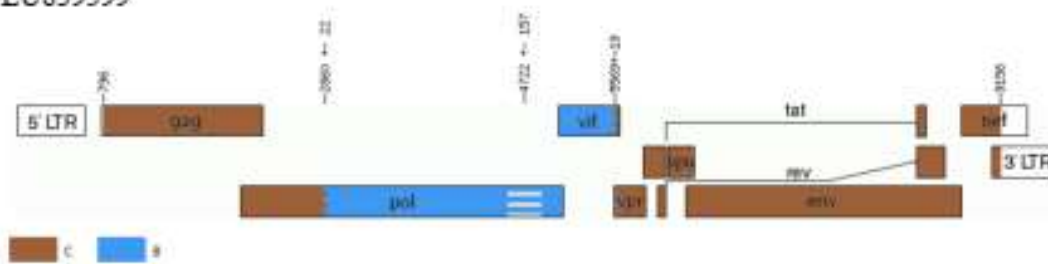
EU000515



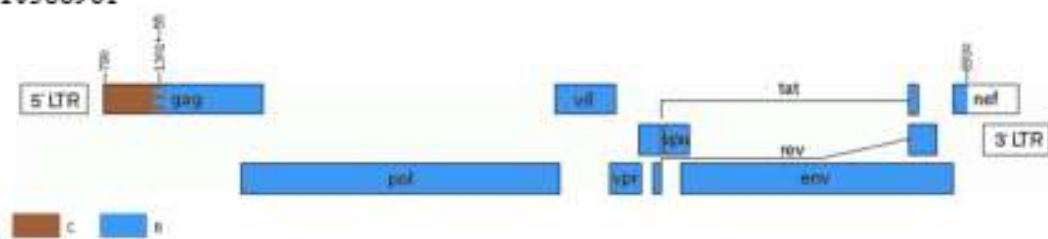
EU000516



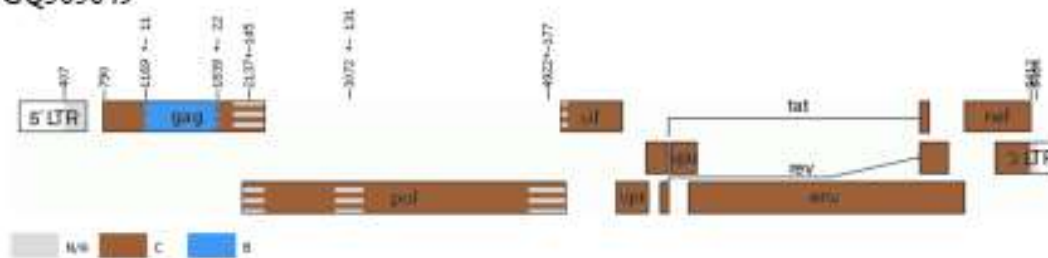
EU839599



FJ388961

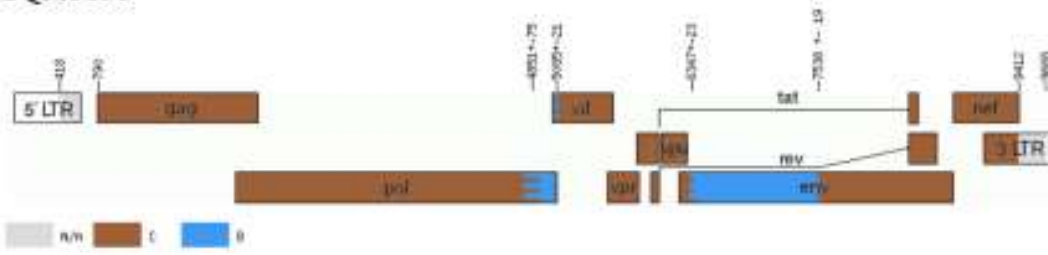


GQ365649

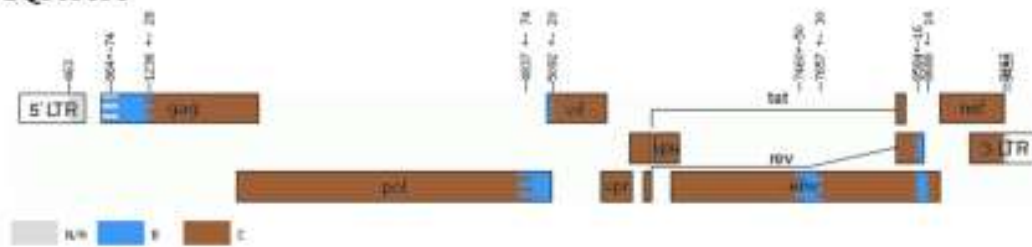




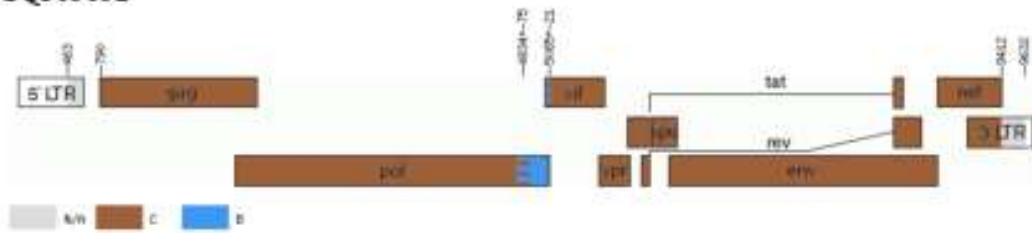
GQ365650



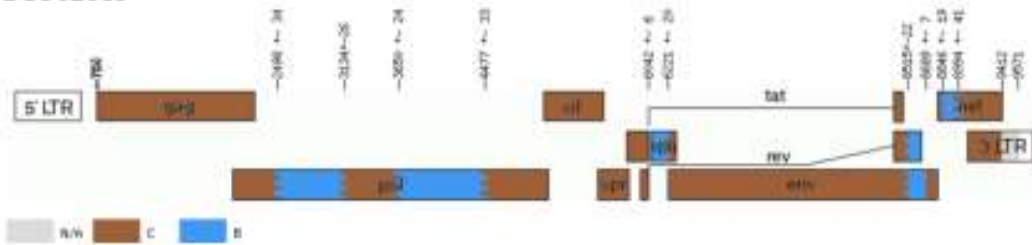
GQ365651



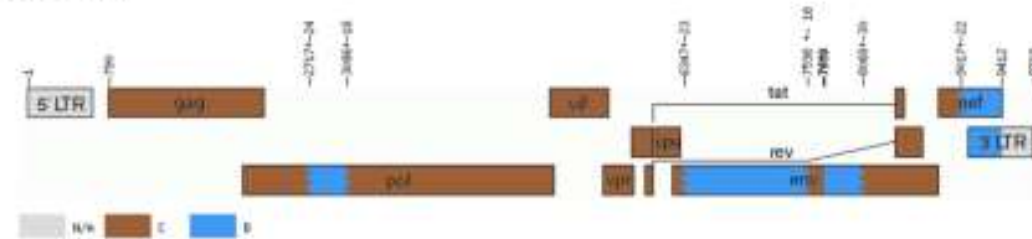
GQ365652



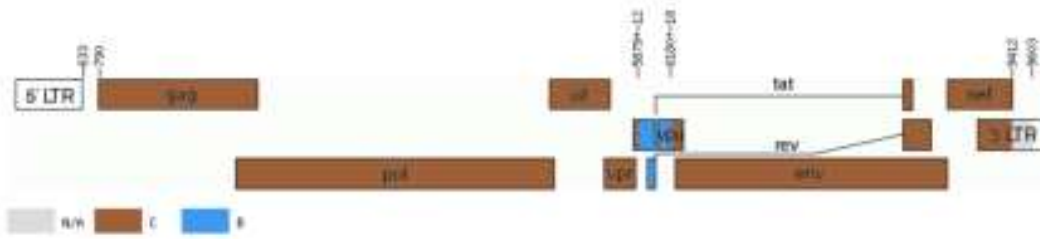
GU362013



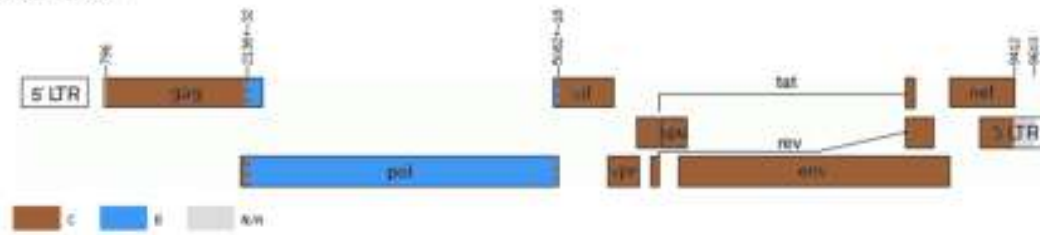
HM100716



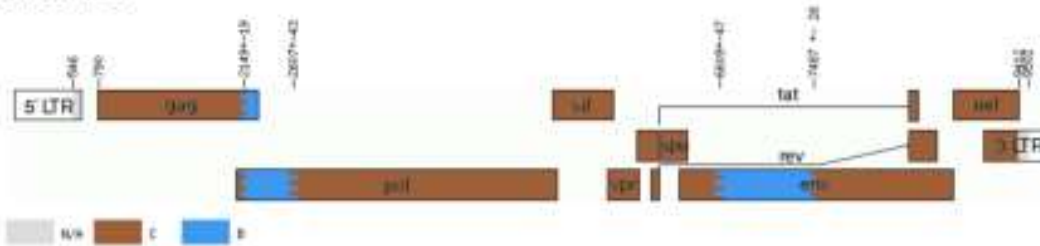
HM776938



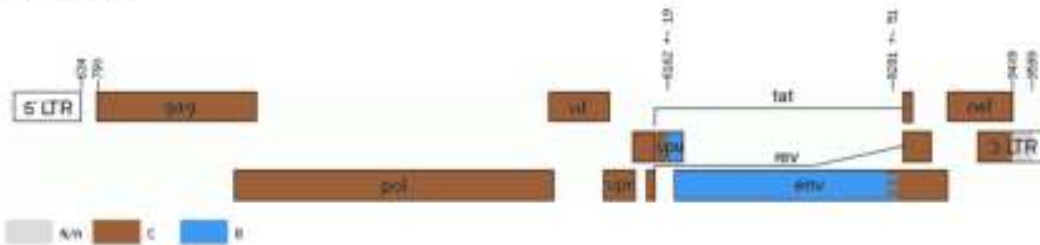
JN235957



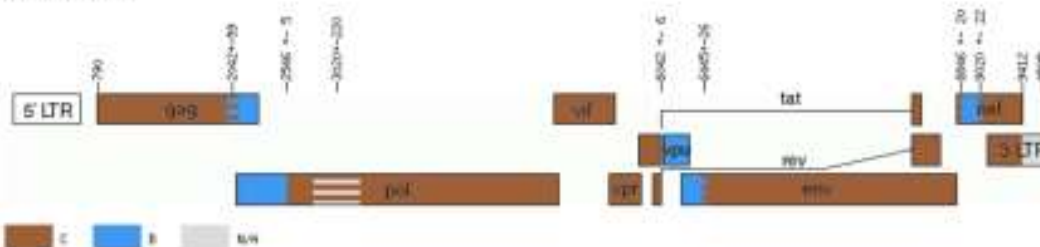
JN692478



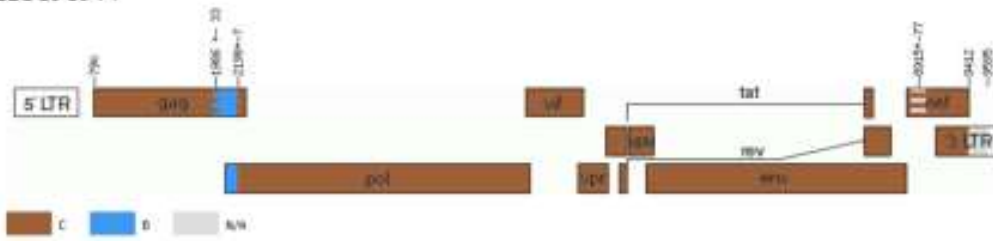
KC596068



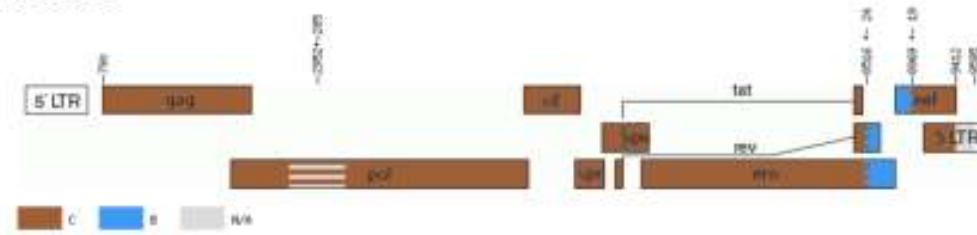
KC898976



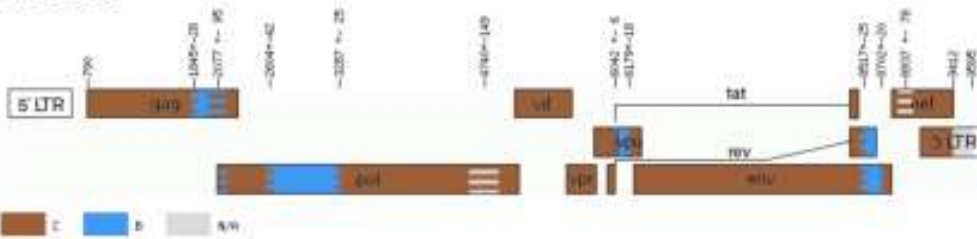
KC898977



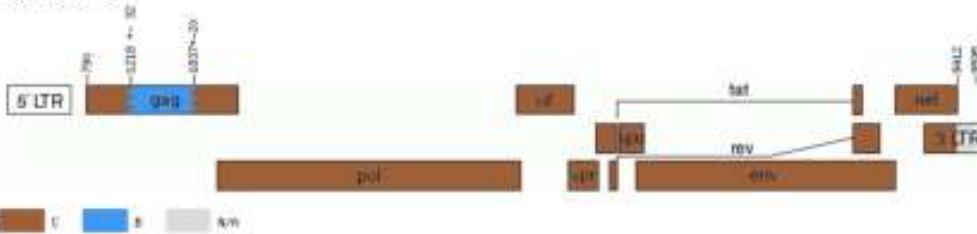
KC898981



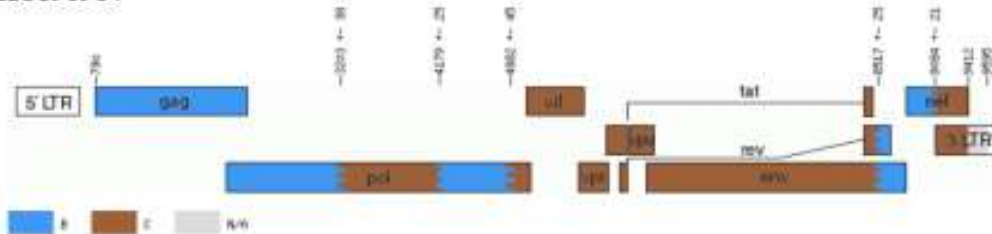
KC898982



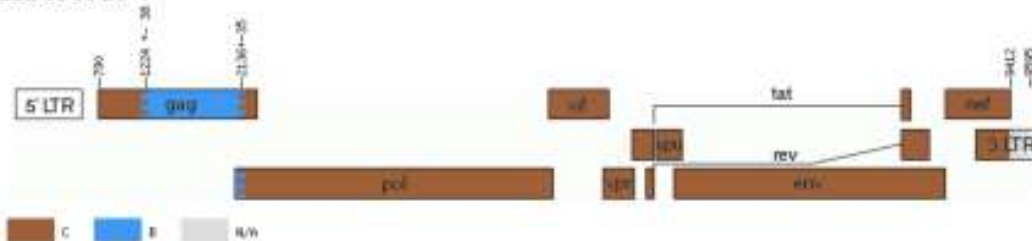
KC898983



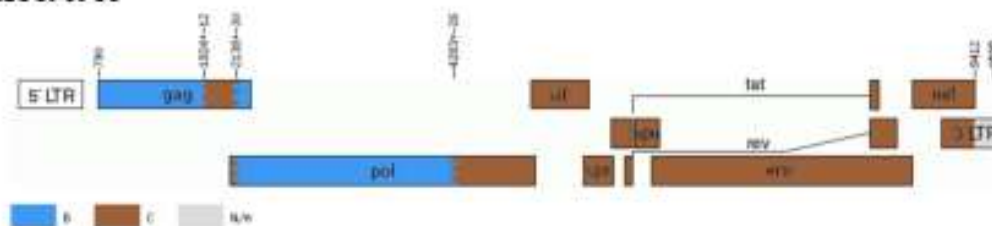
KC898984



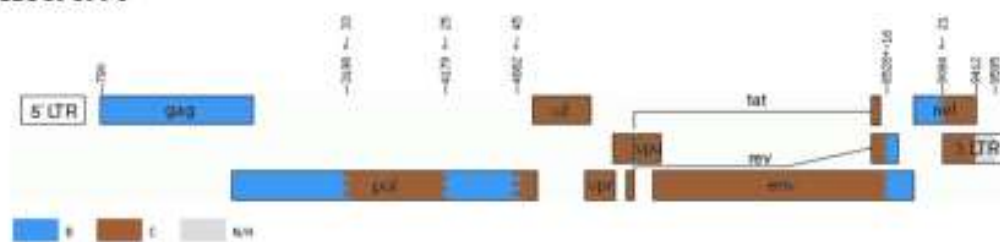
KC898986



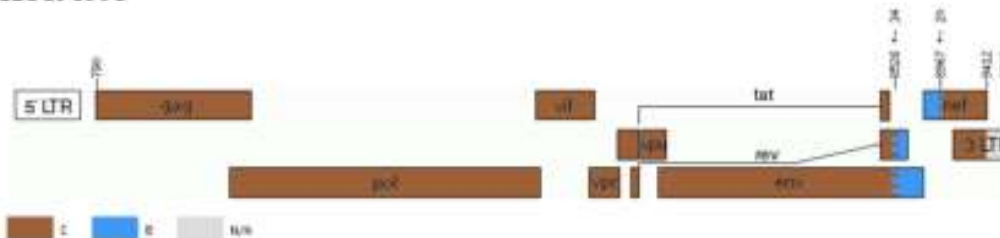
KC898988



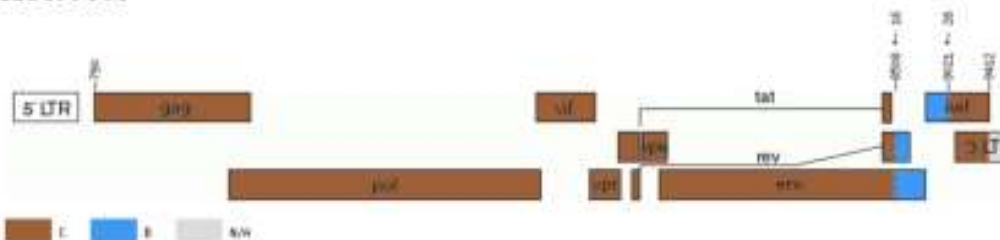
KC898990



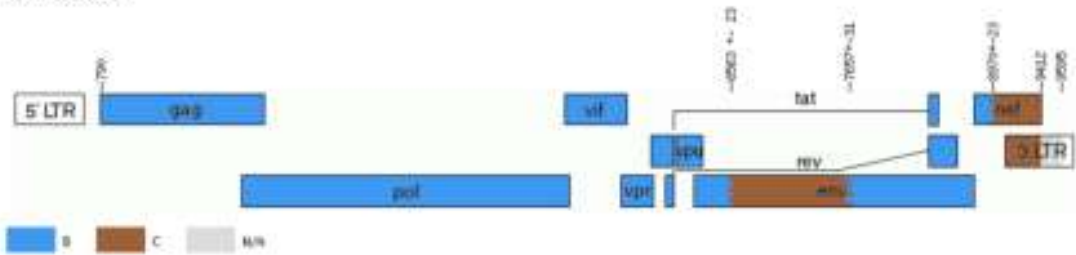
KC898993



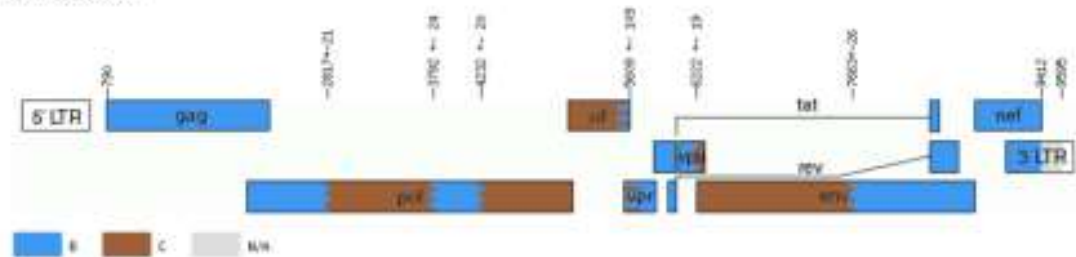
KC899006



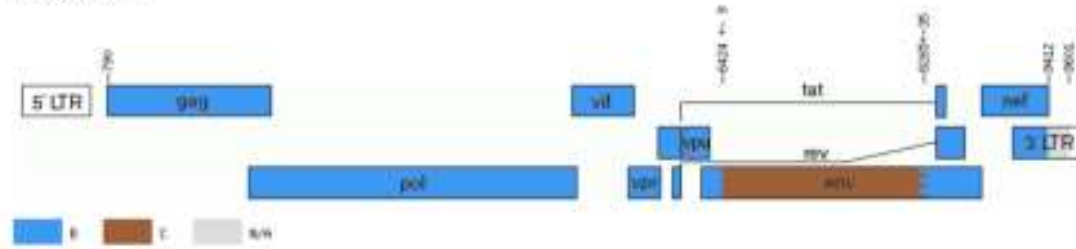
KC899007



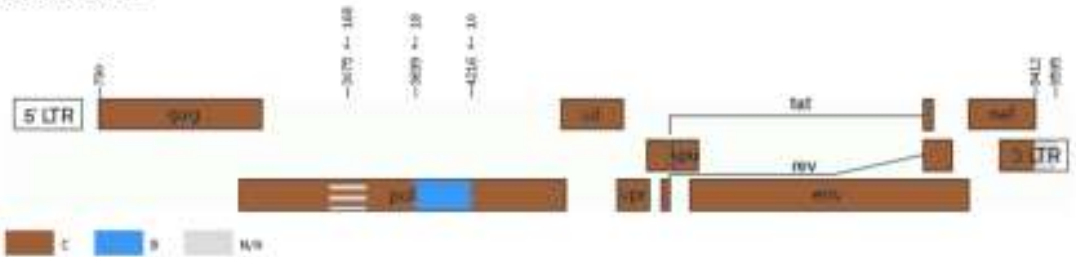
KC899010



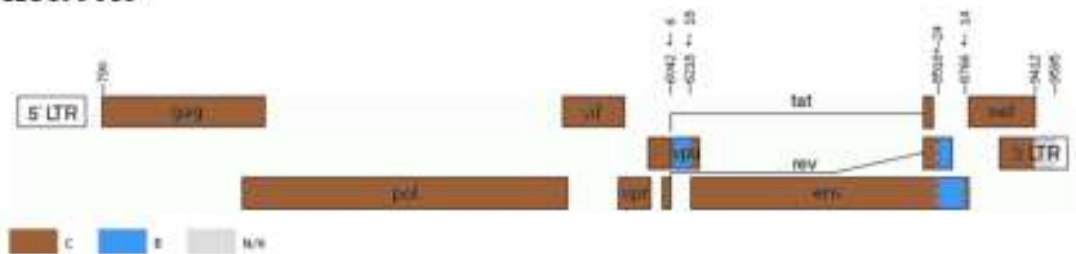
KC899013



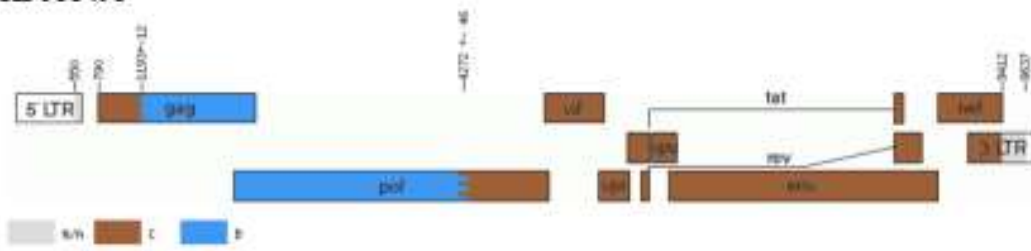
KC899014



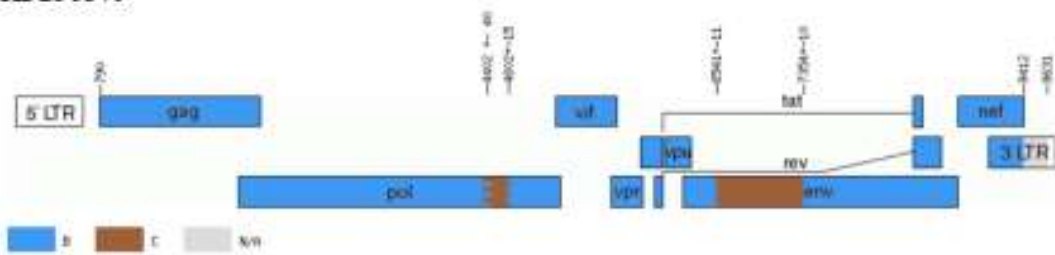
KC899015



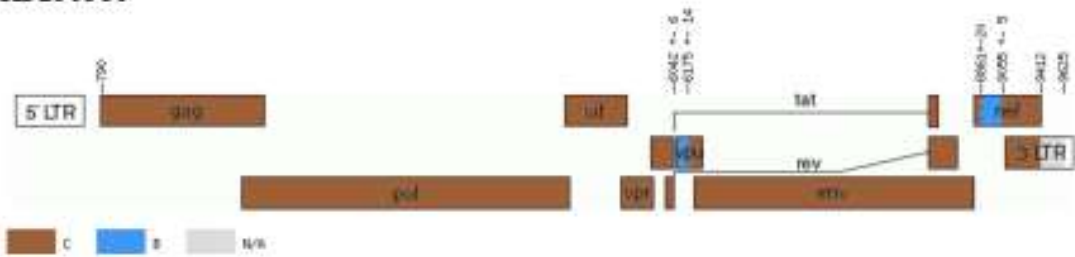
KF011493



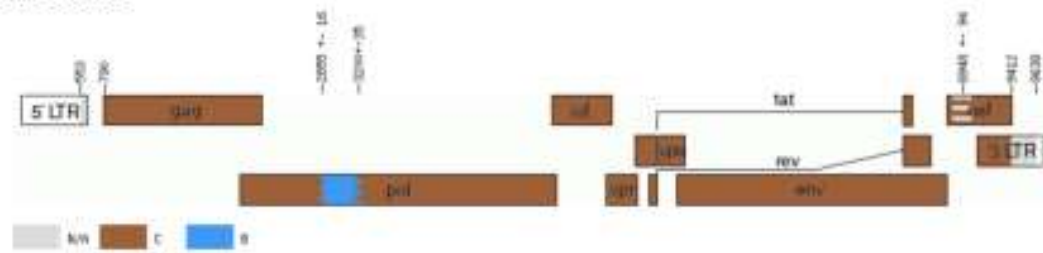
KF250379



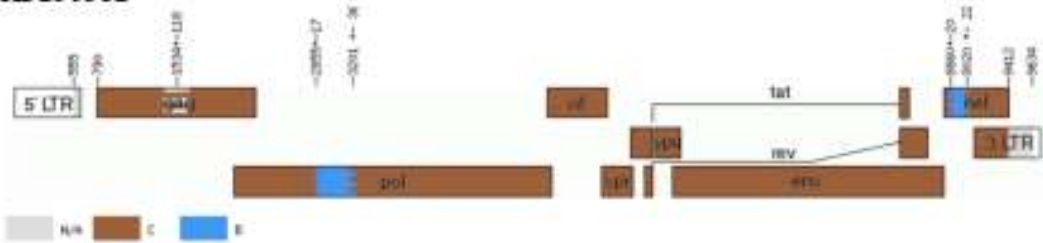
KF250380



KF250381

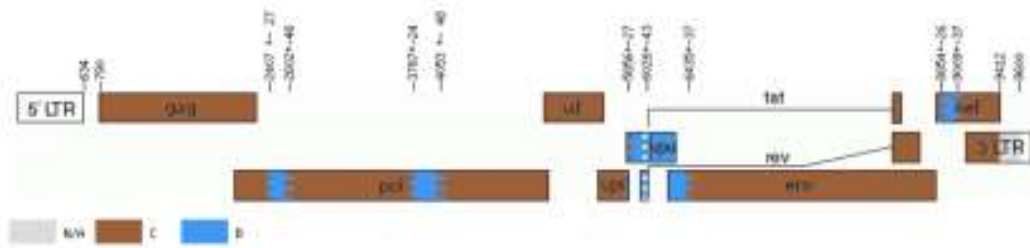


KF250382

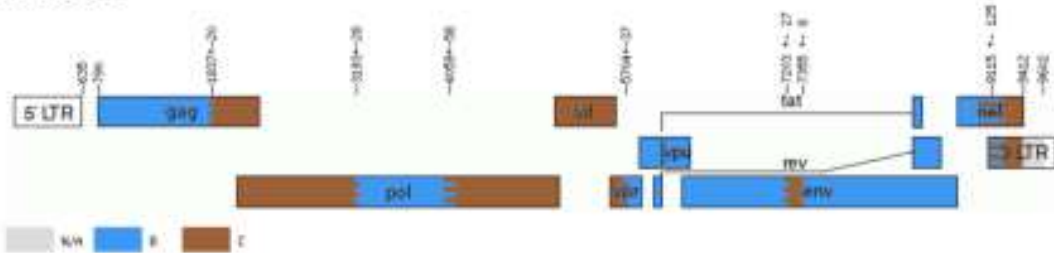




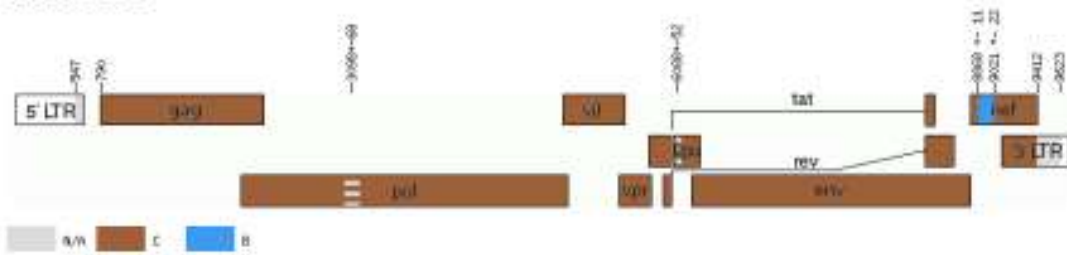
KF250410



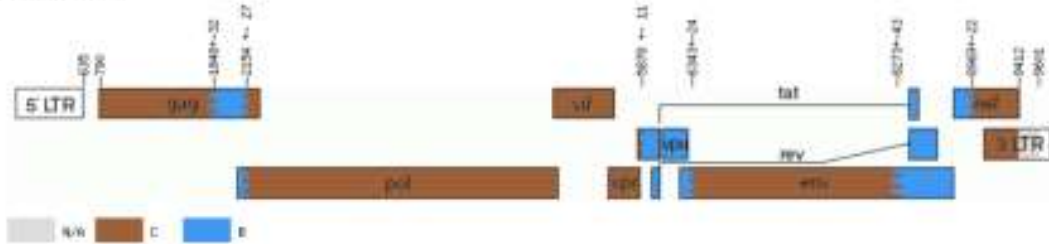
KF835521



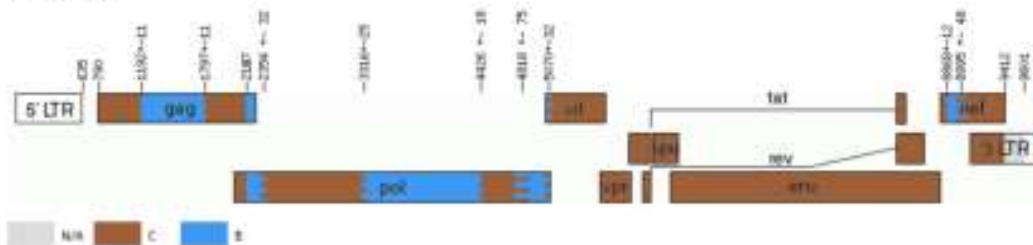
MT635156



KF835525

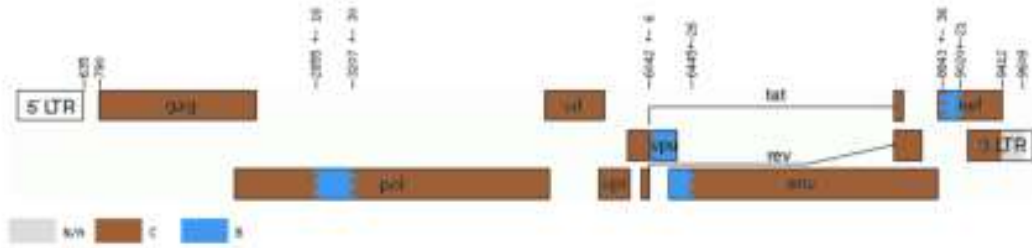


KF835526

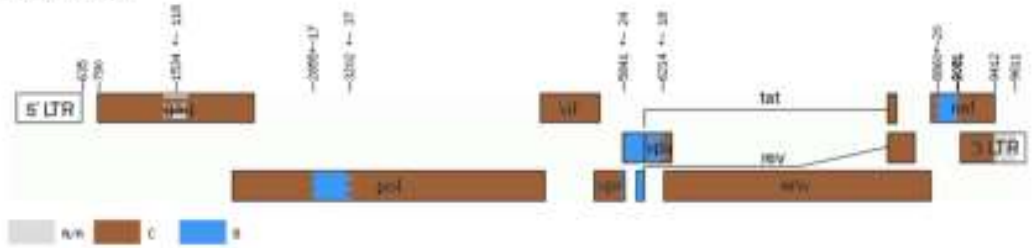




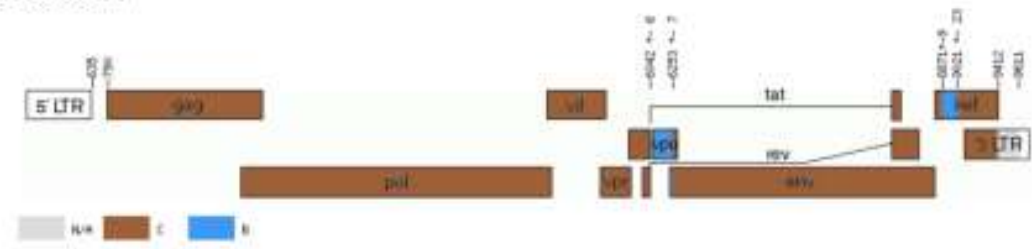
KF835539



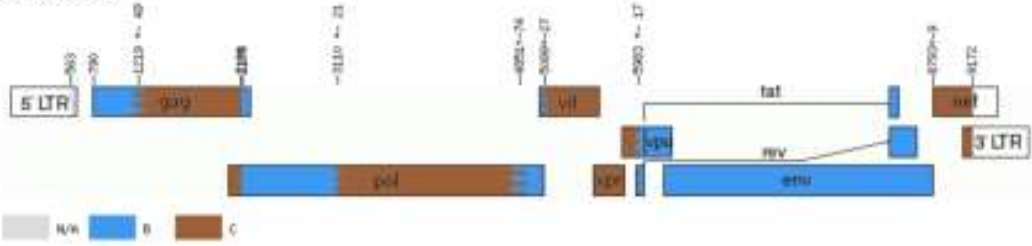
KF835540



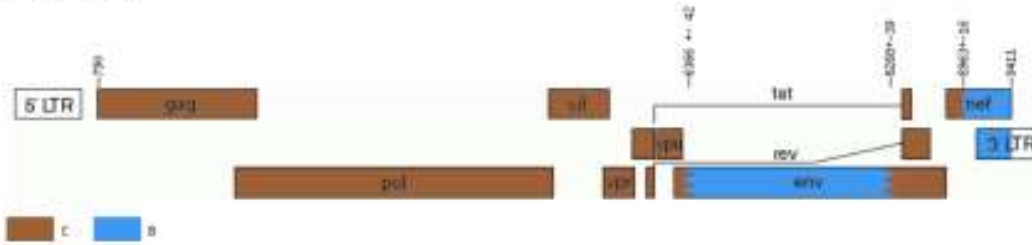
KF835544



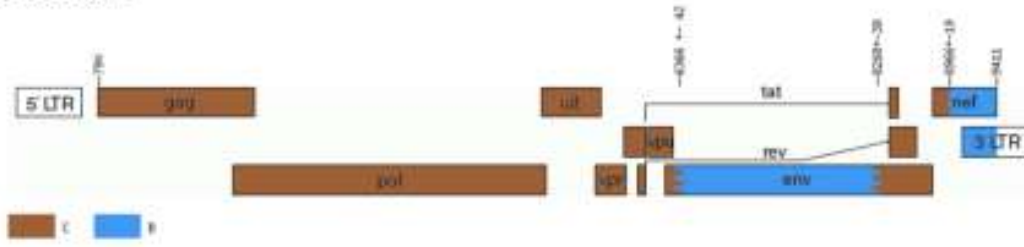
KJ849795



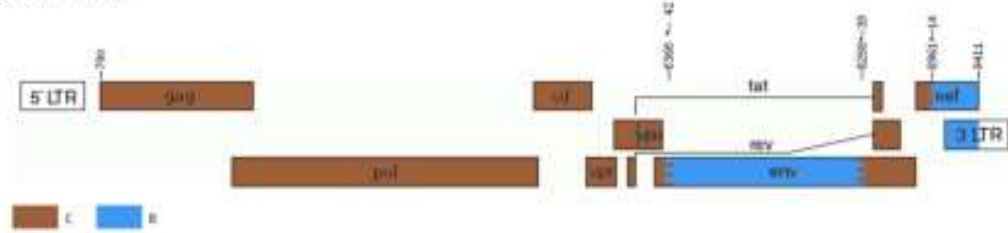
KT183073



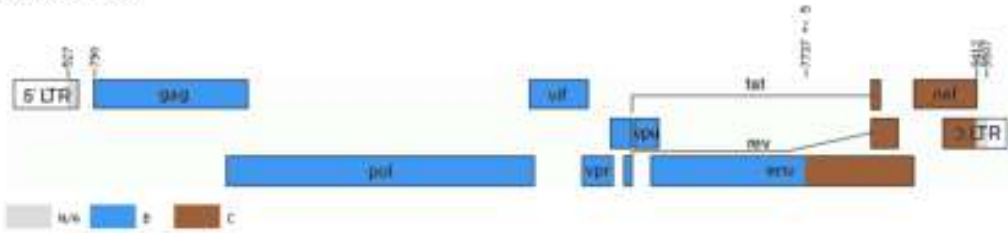
KT183074



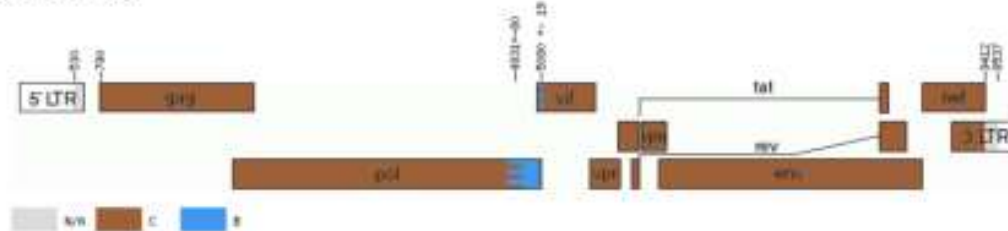
KT183077



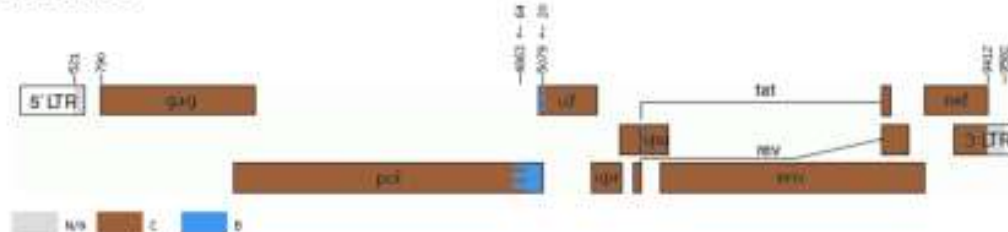
KT427760



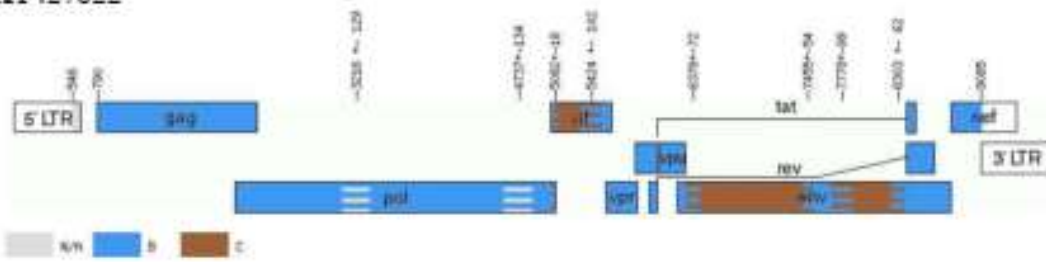
KT427778



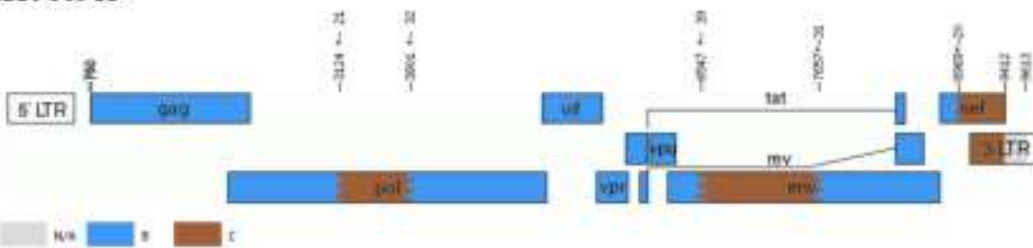
KT427790



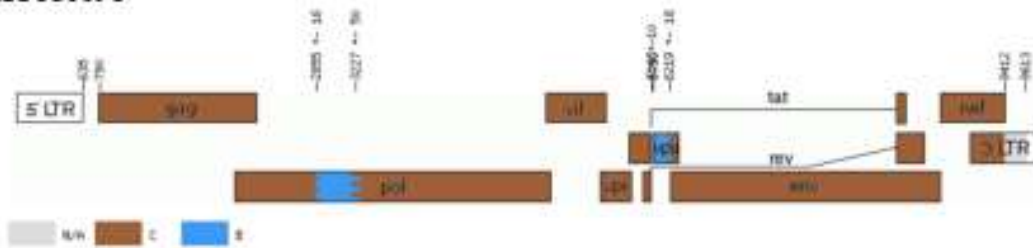
## KT427822



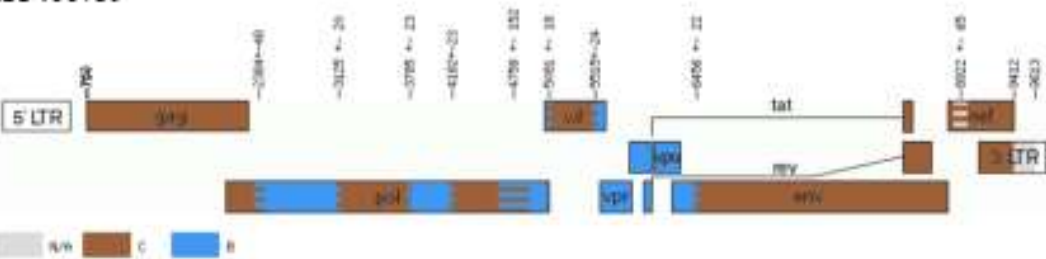
## KT960983



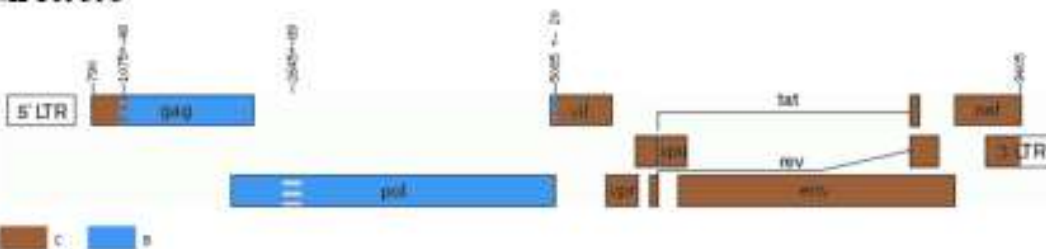
## KU886698



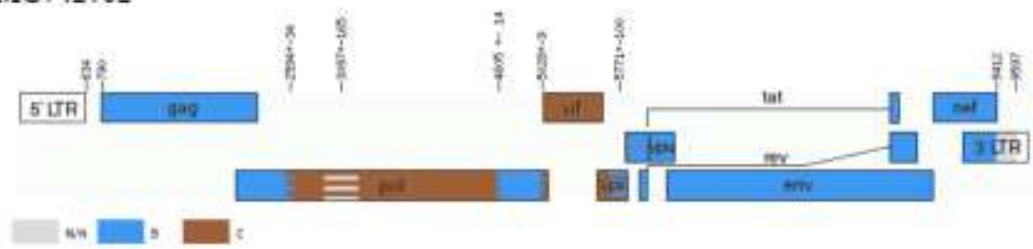
## KY406739



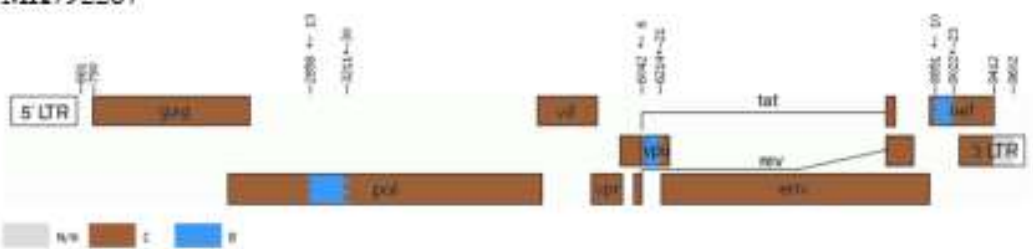
## MF109595



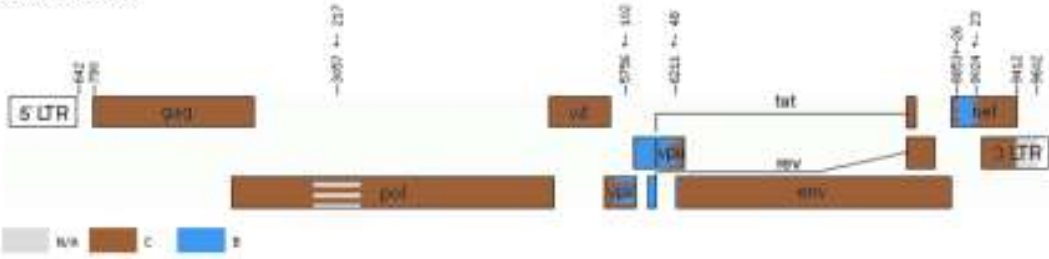
MG742702



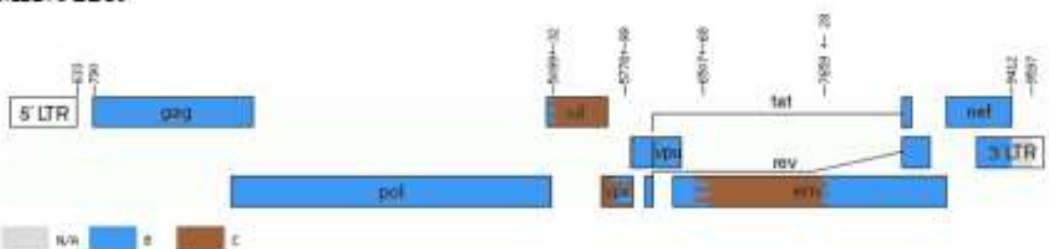
MK792287



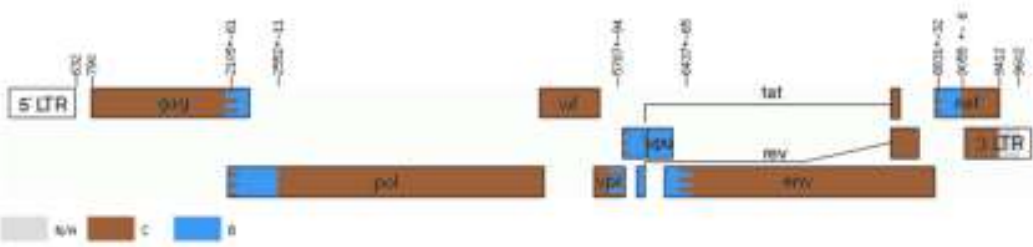
MK792288



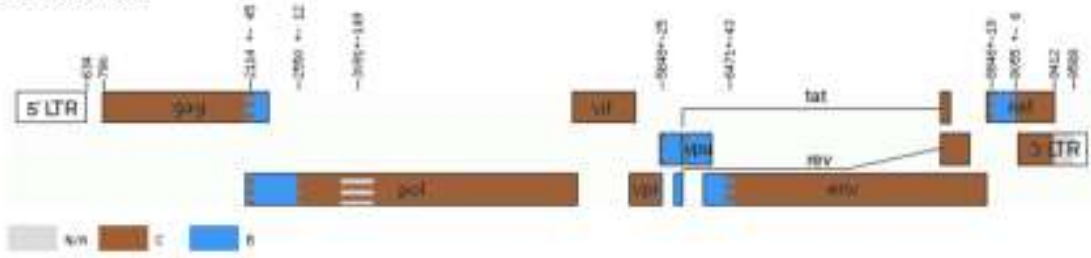
MK792289



MN659380



MN659381



MW145181

