

Universidade de Brasília  
Pós-Graduação em Biologia Molecular

# **História Evolutiva do HIV-1 no Brasil**

Nazle Mendonça Collaço Vêras

Brasília

2010

Universidade de Brasília  
Pós-Graduação em Biologia Molecular

## **História Evolutiva do HIV-1 no Brasil**

Nazle Mendonça Collaço Vêras

Orientador: Luis Fernando de Macedo Brígido (Instituto Adolfo Lutz, São Paulo – SP)

Co-orientador: Marco Salemi (Universidade da Flórida, Gainesville – FL, EUA)

Tese apresentada ao Departamento de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília, como requisito parcial à obtenção do grau de doutor em Biologia Molecular.

Brasília

2010

Universidade de Brasília

## **História Evolutiva do HIV-1 no Brasil**

### Banca Examinadora

Prof<sup>a</sup> Dr<sup>a</sup> Cynthia Kyaw – Universidade de Brasília (Brasília – DF)

Prof Dr Marcelo de Macedo Brígido – Universidade de Brasília (Brasília – DF)

Prof Dr Waldenor Barbosa da Cruz – Universidade de Brasília (Brasília – DF)

Dr<sup>a</sup> Isabel Maria Vicente Guedes de Carvalho Mello – Instituto Butantan (São Paulo - SP)

Dr<sup>a</sup> Alice Nagata – Embrapa Hortaliças (Brasília – DF)

Brasília

2010

*Toda a nossa ciência, comparada  
com a realidade, é primitiva e  
infantil - e, no entanto, é a coisa  
mais preciosa que temos.*

(Albert Einstein)

*Sei que meu trabalho é uma gota  
no oceano. Mas sem ele, o oceano  
seria menor.*

(Madre Tereza de Calcutá)

À minha querida professora e amiga Cláudia Renata Fernandes Martins (*in memoriam*), pelos seus ensinamentos inestimáveis sobre a vida, os quais tenho o privilégio e a honra de carregar comigo todos os dias.

*Soltem as amarras!!! Lancem-se ao  
mar!!! Mas não se esqueçam de  
marcar no mapa o porto de partida!  
Quem sabe, um dia, queiram voltar.  
Se ainda estiver aqui, ficarei feliz em  
saber de suas conquistas e  
orgulhosa quando me contarem  
como superaram as tempestades.*

(Cláudia Renata Fernandes Martins.  
Discurso de patrona da turma de  
Ciências Biológicas da Universidade  
de Brasília em abril de 2002.)

Aos meus pais, Iêda e Adelmir.

À minha irmã, Aira.

À minha avó, Anadir, e tias, Iara e Yete.

Querida família, recebam meu eterno muito obrigada pelo apoio constante e incondicional e por fazerem o impossível, quando necessário, para tornarem meus sonhos realidade.

## Agradecimentos

Ao meu orientador, **Dr Luis Fernando de Macedo Brígido**, por ter me recebido em seu laboratório e nunca ter medido esforços para tornar possível a realização de cada meta planejada. Muitíssimo obrigada por todo o apoio.

Ao meu co-orientador **Prof Dr Marco Salemi**, por toda atenção, paciência e dedicação em me ensinar evolução molecular e a aplicação de análises filogenéticas, de coalescência e filogeográficas no estudo de epidemias virais. Meu muito obrigada pela preciosa colaboração não só na realização deste projeto, como na minha formação profissional.

Aos membros da banca examinadora **Profª Drª Cynthia Kyaw, Prof Dr Marcelo de Macedo Brígido, Prof Dr Waldenor Barbosa da Cruz, Drª Isabel Maria Vicente Guedes de Carvalho Mello e Drª Alice Nagata**, por se disponibilizarem para analisar esse trabalho e pelas valiosas e oportunas observações.

Aos amigos **Daniela Cerqueira, Verônica Veras, Eduardo Ramalho, Claudiner Oliveira, Daniella Moraes, Patrícia Wyant, Regina Amorim, Tainá Alencar, Margareti Medeiros e Natália Milanezi**, do grupo de pesquisa em Virologia da Universidade de Brasília orientado pela Prof Cláudia Renata, por todos os anos de amizade e companheirismo profissional.

À toda equipe do laboratório de genotipagem do Instituto Adolfo Lutz, em especial **Rosângela Rodrigues, João Paulo Batista, Antônio Flávio Siqueira e João Leandro Ferreira**, pelo apoio na realização deste projeto.

Aos amigos **Cristina Oliveira e Artur Queiroz** por todo o aprendizado que construímos juntos, por todos os almoços e *happy hours* que renovavam minhas energias e pela maravilhosa amizade que tornou a vida em São Paulo inesquecível.

À amiga e *roommate* **Luciane Santos**, pela amizade e apoio constante na realização deste projeto. Querida, muito obrigada por todos os momentos compartilhados que fizeram o aprendizado sobre filogenia tão mais fácil e divertido.

À **Rebecca Gray** pelas infinitas horas compartilhadas ao lado do computador, discutindo cada resultado alcançado e ajudando a resolver os mais diversos problemas que surgiram ao longo da realização deste projeto. Você será minha eterna *little genius*.

À **Profª Drª Maureen Goodenow** e toda a equipe do seu laboratório na Universidade da Flórida, em especial **Joshua Bunger, Wilton Williams, Amanda Lowe, Kai-Fen Chang, Li Yin, Mark Wallet, Brent Gardner, Steve Pomeroy e**

**Beverly Osborn**, pelo incrível convívio profissional, pelos comentários sempre tão enriquecedores e, acima de tudo, pela amizade.

À **Ana**, funcionária da Secretaria do Departamento de Biologia Celular, por estar sempre disposta a ajudar em tudo com extrema eficiência.

À **Pós-Graduação em Biologia Molecular da Universidade de Brasília**, pelo incentivo e apoio financeiro.

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)**, pelo apoio financeiro.

À todos os **familiares e amigos**, que compartilharam comigo a realização deste sonho.

À **Deus**, por tornar tudo possível e presentear minha vida com pessoas extraordinárias que fizeram essa caminhada inesquecível.

## Sumário

**Resumo, xii**

**Abstract, xiii**

**Lista de Figuras, xiv**

**Lista de Tabelas, xvi**

**Lista de Quadros, xviii**

**Lista de Siglas e Abreviaturas, xix**

### **1. Revisão Bibliográfica, 21**

1.1. Introdução, 21

1.2. Histórico, 21

1.3. O Vírus da Imunodeficiência Humana (HIV), 23

1.4. Variabilidade Genética do HIV, 25

1.5. Origem do HIV, 28

1.6. Distribuição Geográfica do HIV, 30

1.7. Anti-retrovirais e Resistência, 33

1.8. Resposta Imune e Desenvolvimento de Vacinas, 36

1.9. Evolução Molecular, 43

### **2. Objetivos, 47**

2.1. Objetivo Geral, 47

2.2. Objetivos Específicos, 47

### **3. Metodologia, 48**

3.1. Amostragem, 48

3.1.1. Seqüências genômicas, 48

3.1.2. Técnicas de Processamento de Amostras para Seqüenciamento, 51

3.1.2.1. Extração do RNA Viral, 51

3.1.2.2. Transcrição Reversa, 51

3.1.2.3. Reação em Cadeia da Polimerase (PCR), 52

3.1.2.4. Detecção, Quantificação e Purificação do Produto de PCR, 53

3.1.2.5. Reação de Seqüenciamento, 53

3.1.2.6. Purificação da Reação de Seqüenciamento, 54

3.1.2.7. Edição das Seqüências, 54

3.2. Detecção de Mutações Associadas à Resistência e Análise da Susceptibilidade aos Anti-retrovirais, 54

- 3.3. Mapeamento de Verossimilhança, 55
- 3.4. Análises Filogenéticas, 55
- 3.5. Análise de Recombinação, 56
- 3.6. Análise da Presença de Assinaturas na Protease e Transcriptase Reversa, 57
- 3.7. Identificação de Epitopos na Protease e Transcriptase Reversa, 58
- 3.8. Análise de Relógio Molecular, 58
- 3.9. Análises Filogeográficas, 59

#### **4. Resultados, 61**

- 4.1. Detecção de Mutações Associadas à Resistência e Análise da Susceptibilidade aos Anti-retrovirais, 61
- 4.2. Genotipagem, 78
  - 4.2.1. Definição dos Subtipos do HIV-1, 78
  - 4.2.2. Definição das Formas Recombinantes do HIV-1, 79
- 4.3. Análise da Presença de Assinaturas na Protease e Transcriptase Reversa, 88
- 4.4. Identificação de Epitopos na Protease e Transcriptase Reversa, 93
- 4.5. Análise Filogenética dos Subtipos B e C Circulantes no Brasil, 95
  - 4.5.1. Subtipo B, 95
    - 4.5.1.1. Mapeamento de Verossimilhança, 95
    - 4.5.1.2. Origem do Subtipo B no Brasil, 96
  - 4.5.2. Subtipo C, 100
    - 4.5.2.1. Mapeamento de Verossimilhança, 100
    - 4.5.2.2. Origem do Subtipo C no Brasil, 102
- 4.6. História Demográfica, 105
  - 4.6.1. Subtipo B, 105
  - 4.6.2. Subtipo C, 107
- 4.7. Determinação da Taxa Evolutiva e do Tempo do Ancestral Comum Mais Recente, 108
  - 4.7.1. Subtipo B, 108
  - 4.7.2. Subtipo C, 111
- 4.8. Análises Filogeográficas, 113
  - 4.8.1. Subtipo B, 113
  - 4.8.2. Subtipo C, 115

#### **5. Discussão, 119**

- 5.1. Avaliação da Resistência aos Anti-retrovirais, 119
  - 5.1.1. Análise a Resistência aos Inibidores de Protease, 120

5.1.2. Análise a Resistência aos Inibidores de Transcriptase Reversa,	122
5.2. Definição dos Subtipos e Formas Recombinantes do HIV-1,	125
5.3. Análise da Variabilidade Genética da Protease e Transcriptase Reversa,	127
5.4. Avaliação da Presença de Epitopos na Protease e Transcriptase Reversa,	129
5.5. Origem da epidemia do HIV-1 no Brasil,	131
5.5.1. Subtipo B,	131
5.5.2. Subtipo C,	133
5.6. Determinação do Tempo do Ancestral Comum Mais Recente (Tmrca),	134
5.6.1. Subtipo B,	134
5.6.2. Subtipo C,	135
5.7. Análises Filogeográficas,	136
5.7.1. Subtipo B,	136
5.7.2. Subtipo C,	137
<b>6. Conclusões,</b>	<b>142</b>
<b>7. Referências Bibliográficas,</b>	<b>146</b>
<b>Apêndice I,</b>	<b>159</b>
<b>Apêndice II,</b>	<b>190</b>
<b>Apêndice III,</b>	<b>196</b>
<b>Apêndice IV,</b>	<b>198</b>

## Resumo

A caracterização dos diferentes subtipos do HIV-1 prevalentes em uma região geográfica, assim como a compreensão da origem e disseminação desses subtipos, é essencial para definir estratégias de prevenção e intervenção nas epidemias locais de HIV/AIDS. O presente trabalho analisou 895 seqüências de *pol* disponíveis no *GenBank*, visando descrever os principais subtipos do HIV-1 prevalentes no Brasil, o perfil de resistência de linhagens não expostas a anti-retrovirais, a prevalência de assinaturas e epitopos na protease e transcriptase reversa (RT) e reconstruir a origem e os principais padrões de disseminação dos subtipos B e C do HIV-1 no Brasil. O subtipo B (65,6%) foi o mais prevalente, seguido dos subtipos C (14,1%) e F (6,1%). As formas recombinantes representaram 14,1% das seqüências analisadas. Foram encontradas oito mutações primárias na protease que diminuem a susceptibilidade aos inibidores de protease (PI): L24I, D30N, M46I/L, I54L/V, V82L e I84V. Na RT, foram identificadas 16 mutações associada a resistência aos inibidores de transcriptase reversa análogos de nucleosídeos (NRTI), A62V, D67N/G, T69E/N/S/d, K70R, V75I/L/M, F77L, V118I, Q151M e M184V, e 18 que conferem resistência aos inibidores de transcriptase reversa não-análogos de nucleosídeos (NNRTI), V90I, A98G, L100I, K101E/N/Q, K103N/R, V106I/L, V108I, E138A/G/K, V179D/T, Y181C e G190A. A transmissão de linhagens resistentes aos PI, NRTI e NNRTI foi de 1,2%, 1,8% e 2,1%, respectivamente. Quanto à variabilidade genética de *pol*, foram identificadas duas possíveis assinaturas, N37K e R41N, na protease de linhagens do subtipo C do HIV-1, dois epitopos reconhecidos por células T CD4<sup>+</sup> na protease e 18 por células T CD8<sup>+</sup>, sendo cinco na protease e 13 na RT. Desses últimos, KLVDFRELNK, GIPHPAGLK, TVLDVGDAY, NETPGIRYQY, IRYQYNVL e VIYQYMDDL, são altamente recomendados para a produção de protótipos vacinais a serem aplicados no Brasil. A entrada do subtipo B do HIV-1 no Brasil ocorreu no início da década de 70, pelo Rio de Janeiro e Rio Grande do Sul, a partir dos quais as linhagens se dispersaram exponencialmente em direção a outros estados do Nordeste e Centro-Oeste do país. O subtipo C, por sua vez, entrou no Paraná, em meado da década de 70, e, logo após sua introdução, disseminou exponencialmente no Sul do país. A aplicação de análises filogenéticas de alta resolução, análises de coalescência e filogeografia permitiu a caracterização da epidemia do HIV-1 no Brasil, a re-avaliação de hipóteses e a formulação de novas teorias sobre a origem e disseminação dos subtipos B e C no país.

**Abstract**

Characterizing the different HIV-1 subtypes prevalent in a geographic region as well as understanding the origin and dissemination of these subtypes are essential to define prevention and intervention strategies targeting local HIV/AIDS epidemics. The present study analyzed 895 *pol* sequences, available at GenBank, aiming to describe the main HIV-1 subtypes prevalent in Brazil, the resistance profile of naïve lineages, the prevalence of signatures and epitopes in protease and reverse transcriptase (RT), and reconstruct the origin and the main dissemination patterns of HIV-1 subtypes B and C in Brazil. The subtype B (65.6%) was the most prevalent followed by subtypes C (14.1%) and F (6.1%). The recombinant forms represented 14.1% of the analyzed sequences. At protease, eight major mutations which reduce the susceptibility to protease inhibitors (PIs) were found: L24I, D30N, M46I/L, I54L/V, V82L and I84V. At RT, 16 mutations associated to nucleoside reverse transcriptase inhibitors (NRTIs) resistance, A62V, D67N/G, T69E/N/S/d, K70R, V75I/L/M, F77L, V118I, Q151M and M184V, and 18 mutations that confer resistance to the non-nucleoside reverse transcriptase inhibitors (NNRTIs), V90I, A98G, L100I, K101E/N/Q, K103N/R, V106I/L, V108I, E138A/G/K, V179D/T, Y181C and G190A, were identified. The transmission of lineages resistant to PIs, NRTIs and NNRTIs were 1.2%, 1.8% and 2.1%, respectively. Regarding *pol* genetic variability, we identified two possible signatures, N37K e R41N, at HIV-1 subtype C protease, two T CD4<sup>+</sup> epitopes and 18 T CD8<sup>+</sup> epitopes, of which five were at protease and 13 at RT. Concerning these last ones, KLVDFRELNK, GIPHPAGLK, TVLDVGDAY, NETPGIRYQY, IRYQYNVL and VIYQYMDDL are highly recommended to the production of vaccine prototypes to be applied in Brazil. The HIV-1 subtype B introduction in Brazil occurred in the beginning of 1970s, through Rio de Janeiro and Rio Grande do Sul, from where the lineages exponentially spread to Northeastern and Central Brazil. The subtype C, in turn, entered in Paraná, during mid-1970s, and, briefly after its introduction, started to spread exponentially in the South region. The application of high-resolution phylogenetics, coalescent analysis and phylogeography, allowed an in-depth characterization of the HIV-1 epidemic in Brazil, the reevaluation of hypothesis and the formulation of new insights about the origin and dissemination of subtypes B and C in the country.

## Lista de Figuras

- Figura 1.** Número de crianças e adultos infectados por HIV no mundo em 2008, 22
- Figura 2.** Representação da estrutura da partícula viral do HIV-1, 24
- Figura 3.** Representação esquemática da organização genômica do HIV-1 e HIV-2, 25
- Figura 4.** Filogenia dos grupos M, N e O do HIV-1, 27
- Figura 5.** Árvore filogenética representando a relação entre os diferentes vírus da imunodeficiência em símios (SIV) e linhagens do HIV, 29
- Figura 6.** Distribuição geográfica dos subtipos e formas recombinantes circulantes (CRF) do HIV-1 no mundo, 31
- Figura 7.** Distribuição geográfica dos subtipos e formas recombinantes do HIV-1 no Brasil, 32
- Figura 8.** Principais eventos durante a fase aguda da infecção por HIV-1, 39
- Figura 9.** Representação esquemática de diferentes padrões filogeográficos exibidos por retrovírus que infectam humanos, 46
- Figura 10.** Distribuição geográfica e cronológica das 895 seqüências de *pol* analisadas neste estudo, 48
- Figura 11.** Distribuição geográfica e cronológica das 150 seqüências de *gp120 (env)* do subtipo B do HIV-1 analisadas neste estudo, 49
- Figura 12.** Distribuição geográfica e cronológica das 17 seqüências de *p24* e 13 de *gp41 (env)* do subtipo C do HIV-1 analisadas neste estudo, 50
- Figura 13.** Freqüência de mutações primárias e secundárias associadas à resistência aos inibidores de protease em isolados de HIV-1 de pacientes não expostos a tratamento anti-retroviral, do Sudeste e Sul do Brasil, 61
- Figura 14.** Níveis de resistência de isolados do HIV-1 de pacientes não expostos a anti-retrovirais aos inibidores de protease no Sudeste e Sul do Brasil, 62
- Figura 15.** Freqüência de mutações associadas à resistência aos inibidores de transcriptase reversa análogo de nucleosídeos em isolados de HIV-1 de pacientes não expostos a anti-retrovirais, do Sudeste e Sul do Brasil, 67
- Figura 16.** Níveis de resistência de isolados do HIV-1, de pacientes não expostos a anti-retrovirais, aos inibidores de transcriptase reversa análogos de nucleosídeos no Sudeste e Sul do Brasil, 68
- Figura 17.** Freqüência de mutações associadas à resistência aos inibidores de transcriptase reversa não-análogo de nucleosídeos em isolados de HIV-1 de pacientes não expostos a anti-retrovirais, do Sudeste e Sul do Brasil, 72

**Figura 18.** Níveis de resistência de isolados do HIV-1, de pacientes não expostos a anti-retrovirais, aos inibidores de transcriptase reversa não-análogos de nucleosídeos no Sudeste e Sul do Brasil, 73

**Figura 19.** Prevalência dos subtipos e formas recombinantes do HIV-1 isolado de indivíduos não expostos a tratamento anti-retroviral, do Sudeste e Sul do Brasil, 78

**Figura 20.** Prevalência dos subtipos e formas recombinantes do HIV-1 isolado de indivíduos não expostos a tratamento anti-retroviral, nos diferentes estados do Sudeste e Sul do Brasil, 78

**Figura 21.** Representação esquemática das formas recombinantes do HIV-1 descritas no Sudeste e Sul do Brasil, baseada em *pol*, de acordo com o caracterizado pelo programa *Recombination Detection Program* – RDP, 81

**Figura 22.** Mapeamento de verossimilhança de linhagens do subtipo B do HIV-1 com distribuição global (triângulos superiores) e restrita ao Brasil (triângulos inferiores), 96

**Figura 23.** Árvore filogenética de máxima verossimilhança do subtipo B do HIV-1 baseada no gene da transcriptase reversa, 97

**Figura 24.** Árvore filogenética de máxima verossimilhança do subtipo B do HIV-1 baseada no gene de *gp120 (env)*, 99

**Figura 25.** Mapeamento de verossimilhança de linhagens do subtipo C do HIV-1 com distribuição global (triângulos superiores) e restrita ao Brasil (triângulos inferiores), 101

**Figura 26.** Árvore filogenética de máxima verossimilhança do subtipo C do HIV-1 baseada nos genes *p24 (A)*, da transcriptase reversa (*B*) e *gp41 (env) (C)*, 103

**Figura 27.** Padrões filogeográficos do subtipo B do HIV-1 no Brasil, 113

**Figura 28.** Padrões de migração do subtipo B do HIV-1 no Brasil, 114

**Figura 29.** Rede de fluxo gênico do subtipo B do HIV-1 entre os diferentes estados brasileiros, 115

**Figura 30.** Padrões filogeográficos do subtipo C do HIV-1 no Brasil, 116

**Figura 31.** Mapa de acessibilidade e padrões de migração do subtipo C do HIV-1 no Brasil, 117

**Figura 32.** Rede de fluxo gênico do subtipo C do HIV-1 entre os diferentes estados brasileiros, 118

**Figura 33.** Análise filogenética de máxima verossimilhança de linhagens do subtipo C do HIV-1, incluindo seqüências do Reino Unido, 140

**Figura 34.** Migração do subtipo C do HIV-1 entre Brasil, Reino Unido e África, 141

## Lista de Tabelas

**Tabela 1.** Lista de iniciadores utilizados para amplificar e seqüenciar os genes da protease e transcriptase reversa, 53

**Tabela 2.** Prevalência das assinaturas encontradas da protease (PR) e transcriptase reversa (RT) nas seqüências do subtipo B, C e F circulantes no Sudeste e Sul do Brasil, 89

**Tabela 3.** Identificação das assinaturas encontradas da protease (PR) e transcriptase reversa (RT) nas seqüências do subtipo B, C e F circulantes no Sudeste e Sul do Brasil, nas formas recombinantes caracterizadas por análise filogenética de máxima verossimilhança e pelo programa *Recombination Detection Program* – RDP, 89

**Tabela 4.** Caracterização dos epitopos descritos na protease e transcriptase reversa do HIV-1 circulante no Sudeste e Sul do Brasil, 94

**Tabela 5.** Fator de Bayes entre relógio molecular estrito (RE) e relaxado (RR) aplicados ao alinhamento de *gp120 (env)* dos clados 1 e 2 do subtipo B do HIV-1 circulante no Brasil, 106

**Tabela 6.** Fator de Bayes entre os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana aplicados ao alinhamento de *gp120 (env)* dos clados 1 e 2 do subtipo B do HIV-1 circulante no Brasil, 106

**Tabela 7.** Fator de Bayes entre relógio molecular estrito (RE) e relaxado (RR) aplicados aos alinhamentos de *p24*, transcriptase reversa (RT) e *gp41 (env)* do subtipo C do HIV-1 circulante no Brasil, 107

**Tabela 8.** Fator de Bayes entre os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana aplicados aos alinhamentos de *p24*, transcriptase reversa (RT) e *gp41 (env)* do subtipo C do HIV-1 circulante no Brasil, 108

**Tabela 9.** Tempo do ancestral comum mais recente (Tmrca) e taxa evolutiva do subtipo B aplicando relógio molecular relaxado e os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana, 110

**Tabela 10.** Tempo do ancestral comum mais recente (Tmrca) e taxa evolutiva do subtipo C aplicando relógio molecular relaxado e os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana, 112

**Tabela 11.** Análise de metapopulação do subtipo B do HIV-1 circulante no Brasil, 114

**Tabela 12.** Análise de metapopulação do subtipo C do HIV-1 circulante no Brasil, 116

**Lista de Quadros**

**Quadro 1.** Descrição das seqüências que apresentaram mutações primárias e/ou secundárias na protease relacionando-os ao ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos inibidores de protease, 63

**Quadro 2.** Descrição das seqüências que apresentaram mutações na transcriptase reversa que diminuem a susceptibilidade aos inibidores de transcriptase reversa análogos de nucleosídeos (NRTI) relacionando-os ao ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos NRTI, 69

**Quadro 3.** Descrição das seqüências que apresentaram mutações na transcriptase reversa que diminuem a susceptibilidade aos inibidores de transcriptase reversa não-análogos de nucleosídeos (NNRTI) relacionando-os ao ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos NNRTI, 74

## Lista de Siglas e de Abreviaturas

3TC – lamivudina	HBV – vírus da hepatite B
aa – aminoácidos	HCV – vírus da hepatite C
ABC – abacavir	HIV – vírus da imunodeficiência humana
AIDS – síndrome da imunodeficiência adquirida	HIV-1 – vírus da imunodeficiência humana tipo 1
APV – amprenavir	HIV-2 – vírus da imunodeficiência humana tipo 2
ARV – retrovirus associado à AIDS.	HLA – antígeno de leucócitos humanos
ATV – atazanavir	HMA – ensaio de mobilidade de heteroduplex
AZT – zidovudina	HTLV – vírus da leucemia de células T humana
BSP – Bayesian skyline plot	HTLV-III – vírus T linfotrópico humano do tipo III
cDNA – DNA complementar	ICTV – Comitê Internacional de Taxonomia de Vírus
CRF – formas recombinantes circulantes	IDV – indinavir
CRI – inibidor de co-receptor	II – inibidor de integrase
CTL – linfócito T citotóxico	IN – Integrase
d4T – estavudina	kb – quilobase
DC – células dendríticas	LAV – vírus associado a linfadenopatia
ddC – zalcitabina	LPV – lopinavir
ddI – didanosina	LTR – seqüências terminais repetidas longas
DLV – delavirdina	MAP – máxima probabilidade posterior
DNA – ácido desoxirribonucléico	MCMC – Monte Carlo de cadeias de Markov bayesianas
DRV – duranavir	MHC – complexo maior de histocompatibilidade
EFV – efavirenz	mL- mililitro
ENF ou T20 – enfuvirtide	
ETR – etravirina	
FB – fator de Bayes	
FI – inibidor de fusão	
FPV – fosamprenavir	
FTC – emtricitabina	
g - gravidade	
GALT – tecido linfóide associado ao intestino	
HAART – terapia anti-retroviral altamente ativa	

MV – máxima verossimilhança

NFV – nelfinavir

ng – nanograma

NK – células natural killer

nm – nanômetros

NNRTI – inibidor de transcriptase reversa não-análogo de nucleotídeo

NRTI – inibidor de transcriptase reversa análogo de nucleotídeo

NVP – nevirapina

ORF – fase de leitura aberta (open reading frame)

pb – pares de base

PI – inibidor de protease

PR – protease

RENAGENO – Rede Nacional de Genotipagem

RMP – reconstrução mais parsimônia

RNA – ácido ribonucléico

RT – transcriptase reversa

RTV – ritonavir

SIV – vírus da imunodeficiência em símios

SQV – saquinavir

TAE – tamanho amostral efetivo

TAM – mutações associadas a análogos de timidina

TDF – tenofovir

TLR – receptor tool-like

Tmrca – tempo do ancestral comum mais recente

TPV – tipranavir

UTH – unidade taxonômica hipotética

UTO – unidade taxonômica operacional

v – volt

WHO – Organização Mundial de Saúde

µg – micrograma

µL – microlitro

## 1. Revisão Bibliográfica

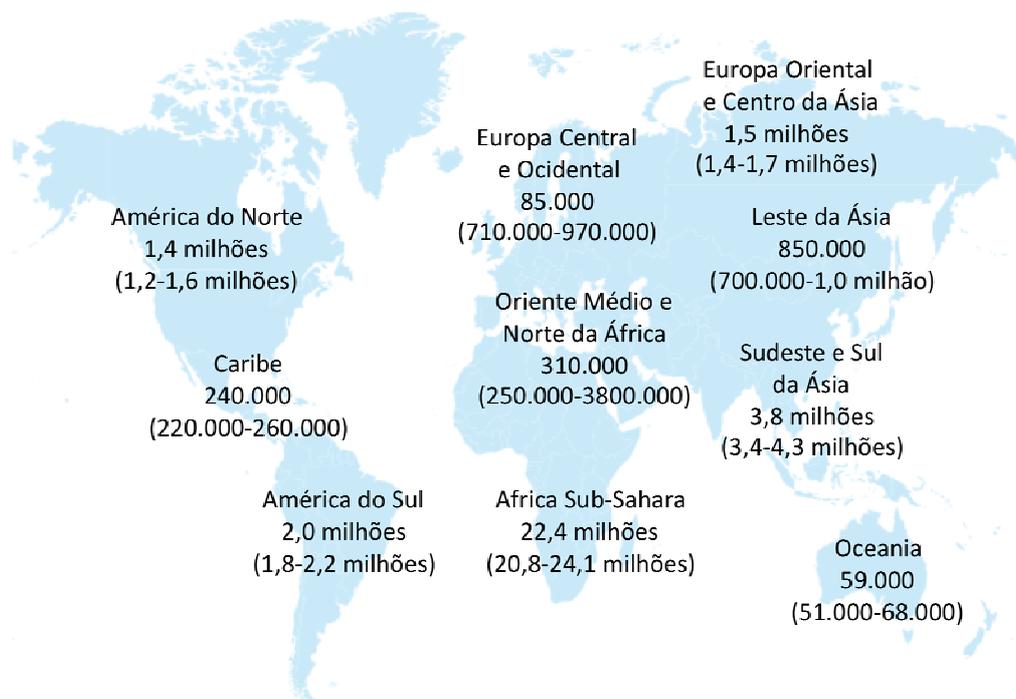
### 1.1. Introdução

A síndrome da imunodeficiência adquirida (*acquired immune deficiency syndrome* – AIDS) é uma das maiores pandemias da história da humanidade. Mundialmente, estima-se que, aproximadamente, 35 milhões de pessoas estejam infectadas pelo vírus da imunodeficiência humana (*human immunodeficiency virus* – HIV), havendo 2,7 milhões de novas infecções só em 2008 (Unaid/Who, 2009a) (Figura 1).

Na América Latina, a epidemia permanece estável com uma prevalência de infecções por HIV de 0,6%. No entanto, em 2008, ocorreram cerca de 170.000 infecções novas por HIV e estima-se que 2 milhões de indivíduos vivam com o vírus nessa região (Unaid/Who, 2009a). No Brasil, até o final de 2007, cerca de 730 000 indivíduos estavam infectados pelo vírus (Unaid, 2008).

De acordo com a UNAIDS/WHO (Unaid/Who, 2007), um terço da população infectada por HIV na América Latina residem no Brasil. De 1980 a junho de 2008, foram notificados no Brasil 506.499 casos de AIDS. Embora inicialmente concentrada em homens homossexuais, nos últimos anos, a epidemia se alastrou por populações com diferentes comportamentos de risco, principalmente usuários de drogas e mulheres que se contaminaram durante a relação sexual. O Sudeste é a região geográfica que apresenta maior número de casos de AIDS desde a década de 80, com 305.725 notificações. Desde 1996, as maiores taxas de incidência de casos de AIDS (por 100.000 habitantes) ocorrem no Sul e Sudeste (Brasil, 2008a).

Não há dúvida que a AIDS continua sendo um dos problemas de saúde mais sérios no país. Portanto, é extremamente importante identificar a origem do vírus e entender como ocorreu e quais os fatores relacionados à dispersão do HIV no Brasil. Essas informações têm impacto direto em medidas de controle e predição da epidemia.



<b>Número de pessoas infectadas por HIV</b>	<b>33,4 milhões (31,1-35,8 milhões)</b>
<b>Novas infecções por HIV em 2008</b>	<b>2,7 milhões (2,4-3,0 milhões)</b>
<b>Mortes por AIDS em 2008</b>	<b>2,0 milhões (1,7-2,4 milhões)</b>

**Figura 1. Número de crianças e adultos infectados por HIV no mundo em 2008.** (adaptado de (Unaid/Who, 2009b).

## 1.2. Histórico

Os primeiros casos de AIDS foram relatados em 1981, nos Estados Unidos, em adultos homossexuais do sexo masculino, que foram a óbito devido a doenças como sarcoma de Kaposi e pneumonia por *Pneumocytis carinii*, acompanhadas por um acentuado decréscimo na contagem de células T CD4<sup>+</sup> (Centers for Disease Control, 1981). No entanto, estudos retrospectivos isolaram o HIV de uma amostra (ZR59) de 1959 da República Democrática do Congo (Zaire), África (Zhu, Korber *et al.*, 1998).

A primeira indicação que a AIDS poderia ser causada por um retrovírus surgiu em 1983, quando um grupo de pesquisadores liderados pelo Dr. Luc Montagnier, do Instituto Pasteur de Paris, isolou um vírus, contendo a enzima transcriptase reversa, do linfonodo de um paciente com síndrome de linfadenopatia persistente. O isolado foi denominado de vírus associado à linfadenopatia (*lymphadenopathy-associated virus* – LAV) (Barre-Sinoussi, Chermann *et al.*, 1983). O retrovírus descrito apresentava, ainda, algumas características peculiares ao vírus da leucemia de células

T humana (*human T-cell leukemia virus* – HTLV), como seu tropismo por células T auxiliares, o que levou a alguns cientistas acreditarem que o LAV pertencia a esse grupo de retrovírus já conhecidos.

Posteriormente, Dr. Robert Gallo, dos Institutos Nacionais de Saúde dos Estados Unidos (*National Institutes of Health* - NIH), descreveu um outro vírus, isolado de células mononucleares do sangue periférico de pacientes com AIDS, denominando-o de vírus T linfotrópico humano do tipo III (*human T-lymphotropic virus type III* - HTLV-III) (Gallo, Sarin *et al.*, 1983). Em 1984, Levy e colaboradores isolaram um vírus de pacientes com diferentes comportamentos de risco, que manifestavam ou não o quadro de AIDS, e o denominaram de retrovírus associado à AIDS (*AIDS-associated retroviruses* – ARV) (Levy, Hoffman *et al.*, 1984). A identificação do ARV em indivíduos assintomáticos foi a primeira indicação do estado de portador. Em 1986, devido à sinonímia, o Comitê Internacional de Taxonomia de Vírus (*International Committee on Taxonomy of Viruses* – ICTV) recomendou a denominação desse agente de vírus da imunodeficiência humana (*human immunodeficiency virus* – HIV) (Coffin, Haase *et al.*, 1986; International Committee on Taxonomy of Viruses, 2006).

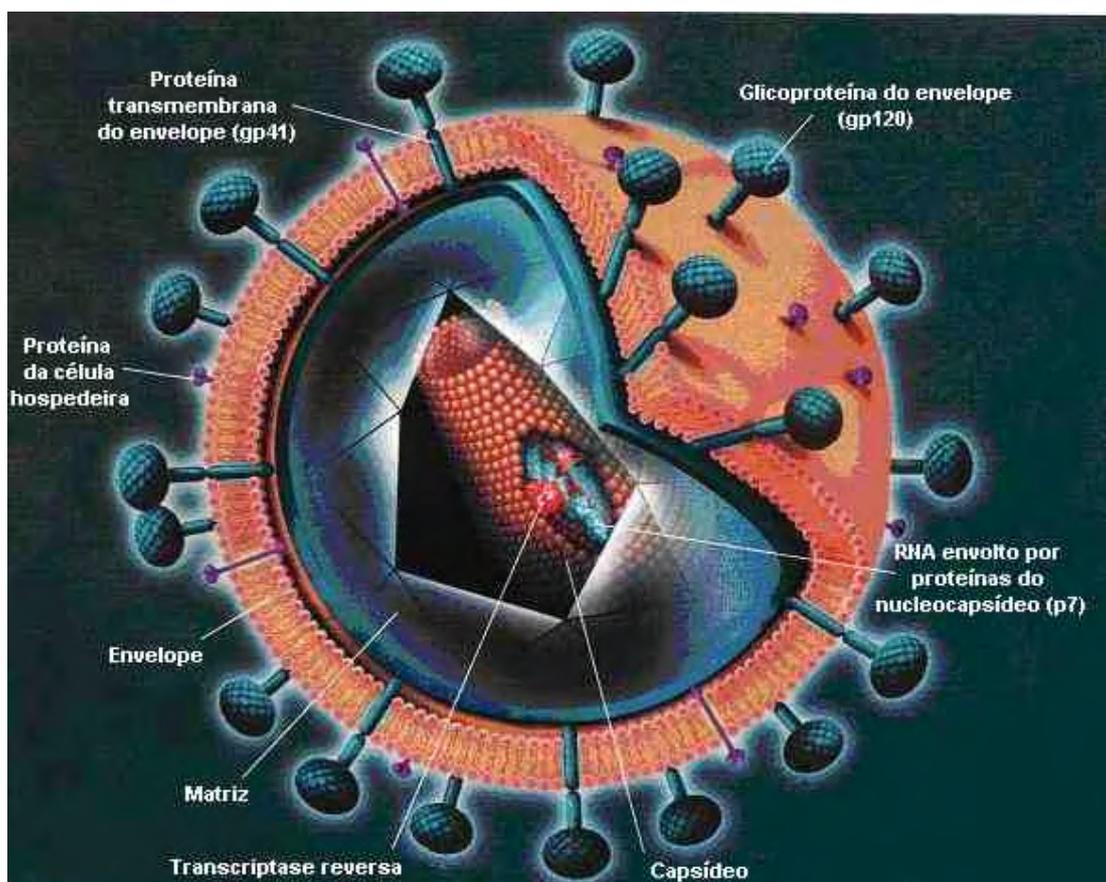
Ainda em 1986, um outro vírus com características semelhantes foi isolado a partir de indivíduos da região ocidental da África, sendo denominado vírus da imunodeficiência humana tipo 2 (*human immunodeficiency virus type 2* – HIV-2). Ambos causam AIDS, mas podem ser distinguidos por suas organizações genômicas e relações filogenéticas com outros lentivirus de primatas. Além disso, a progressão da doença causada pelo HIV-2 é mais lenta (Clavel, Guetard *et al.*, 1986).

### **1.3. O Vírus da Imunodeficiência Humana (HIV)**

O HIV pertence à família *Retroviridae*, gênero *Lentivirus*, cujos membros apresentam genoma formado por duas fitas simples de RNA positivas idênticas, capsídeo em forma de cone e longo período de incubação (Turner e Summers, 1999; International Committee on Taxonomy of Viruses, 2006).

A partícula viral infectiva tem um diâmetro de, aproximadamente, 100nm. O HIV (Figura 2) possui um envelope derivado da membrana plasmática da célula. Nele, além de proteínas celulares, como actina e ubiquitinas, encontram-se glicoproteínas virais de superfície (gp120) ancoradas em proteínas virais transmembranas (gp41). Na face interna do envelope viral, localiza-se a matriz, que é formada por, aproximadamente, 2.000 cópias de proteínas da matriz (p17). O

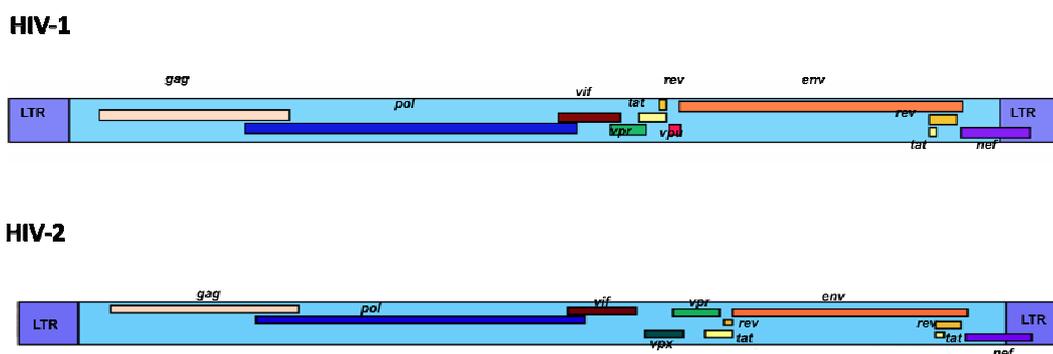
capsídeo, por sua vez, apresenta um formato cônico e ocupa a porção central da partícula viral, sendo constituído por cerca de 2.000 moléculas de proteínas do capsídeo (p24). O genoma viral é composto por duas fitas de RNA positivas, estabilizadas por proteínas do nucleocapsídeo (p7). No interior do capsídeo, encontram-se, ainda, as enzimas protease (PR), transcriptase reversa (RT) e integrase (IN) e as proteínas Nef, Vif e Vpr (Turner e Summers, 1999).



**Figura 2. Representação da estrutura da partícula viral do HIV-1.** O HIV é envolto por um envelope de natureza lipídica. Em sua face interna, localiza-se a matriz e, na porção central da partícula viral, encontra-se o capsídeo com formato cônico. O genoma viral é composto por duas fitas de RNA positivas, estabilizadas por proteínas do nucleocapsídeo (p7). No interior do capsídeo, encontram-se, ainda, as enzimas protease (PR), transcriptase reversa (RT) e integrase (IN) e as proteínas Nef, Vif e Vpr (adaptado de Larder *et al.*, 2001).

O HIV apresenta um genoma de 9,8kb com ORF (*open reading frames*) que codificam diversas proteínas virais. Três genes, *gag*, *pol* e *env*, são comuns a todos os retrovírus e codificam proteínas estruturais. O gene *gag* codifica precursores de proteínas do capsídeo, matriz e nucleocapsídeo, *pol* codifica os precursores das enzimas protease, transcriptase reversa e integrase e *env* codifica precursores das glicoproteínas e proteínas transmembranas do envoltório. Além desses, existem dois

genes regulatórios, *tat* e *rev*, responsáveis pelo controle da expressão gênica em nível transcricional e pós-transcricional, respectivamente, e quatro acessórios, *nef*, *vif*, *vpr* e *vpu*, responsáveis pela modulação do ciclo viral. A diferença genômica entre HIV-1 e HIV-2 reside na presença do gene *vpu* no HIV-1 e de *vpx* no HIV-2. Nas terminações 3' e 5', há longas seqüências repetidas, chamadas LTR (*long terminal repeat sequences*), que contêm seqüências necessárias para a transcrição reversa, integração, início e término da transcrição do genoma viral integrado ao DNA celular, e ligação da proteína reguladora Tat (Figura 3) (Turner e Summers, 1999).



**Figura 3. Representação esquemática da organização genômica do HIV-1 e HIV-2.** As posições verticais denotam as diferentes fases de leitura que codificam as proteínas virais. Os genes *gag*, *pol* e *env* codificam proteínas estruturais. Além desses, existem dois genes regulatórios, *tat* e *rev*, e quatro acessórios, *nef*, *vif*, *vpr* e *vpu*. A diferença genômica entre HIV-1 e HIV-2 reside na presença do gene *vpu* no HIV-1 e de *vpx* no HIV-2. Nas terminações 3' e 5', há longas seqüências repetidas, chamadas LTR (adaptado de (Flint, Enquist *et al.*, 2004).

#### 1.4. Variabilidade Genética do HIV

O HIV apresenta uma alta variabilidade genética, resultante de taxas elevadas de geração de novos vírions (10 bilhões de partículas virais por dia, durante a infecção) e de substituições nucleosídicas incorretas durante a transcrição reversa (aproximadamente  $10^{-4}$  por nucleotídeo) que, por sua vez, levam a mutações ( $3 \times 10^{-5}$  mutações por nucleotídeo por dia) (Morgado, Guimaraes *et al.*, 2002; Thomson, Perez-Alvarez *et al.*, 2002; Kijak e Mccutchan, 2005; Requejo, 2006).

A diversidade genética do HIV-1 em uma pessoa infectada é expressa por subpopulações de variantes virais geneticamente distintos, porém relacionados, denominados quasispécies. A distância genética entre as quasispécies é descrita como divergência genética. A geração de diversidade e divergência intra-quasispécies facilita a evolução do HIV-1 durante a progressão da doença (Kijak e Mccutchan, 2005; Mccutchan, 2006; Requejo, 2006).

Através de estudos sorológicos, evidenciaram-se, até o momento, dois tipos antigênicos: HIV-1 e HIV-2. O HIV-1 é o tipo mais virulento e amplamente disseminado no mundo. Já o HIV-2 é menos transmissível e a progressão da doença é mais lenta, quando comparado ao HIV-1. O HIV-2 é encontrado, quase que exclusivamente, no Oeste da África (Turner e Summers, 1999; Kijak e Mccutchan, 2005; Requejo, 2006).

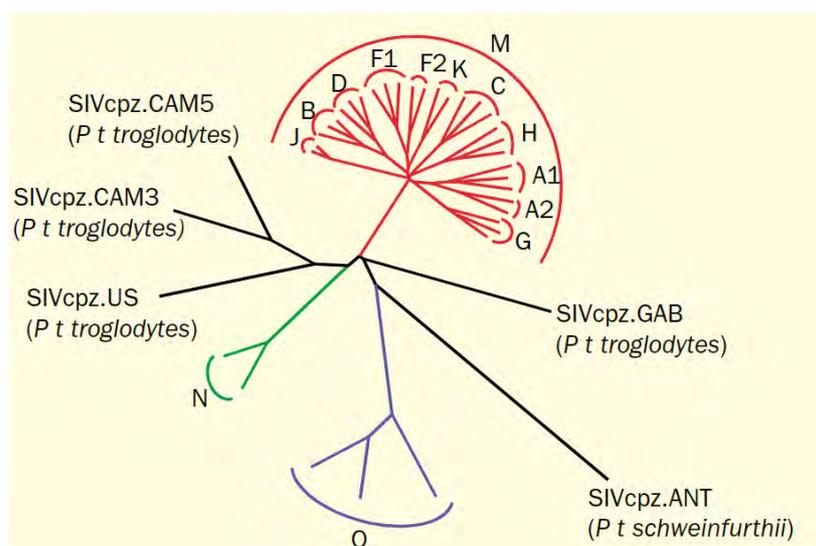
De acordo com análises filogenéticas de várias linhagens de HIV-1, isoladas de diversas regiões geográficas, esse vírus pode ser classificado em grupos, subtipos, sub-subtipos e formas recombinantes circulantes (Robertson, Anderson *et al.*, 1999; Turner e Summers, 1999; Peeters, 2000; Morgado, Guimaraes *et al.*, 2002; Takeb, Kusagawa *et al.*, 2004; Kijak e Mccutchan, 2005; Mccutchan, 2006; Requejo, 2006)

Foram descritos três grupos distintos: M (*major*), O (*outlier*) e N (*new* ou *non-M non-O*). O grupo M é composto pela maioria das linhagens do HIV-1 responsáveis pela pandemia da AIDS. O grupo O é representado por linhagens altamente divergentes, endêmicas na República de Camarões e em alguns países vizinhos da África Central Ocidental, além de alguns casos isolados na Europa e nos Estados Unidos. O grupo N, identificado recentemente, é constituído por um número limitado de linhagens isoladas na República de Camarões, altamente relacionadas ao SIV<sub>cpz</sub> de chipanzés da região (Robertson, Anderson *et al.*, 1999; Turner e Summers, 1999; Peeters, 2000; Morgado, Guimaraes *et al.*, 2002; Takeb, Kusagawa *et al.*, 2004; Kijak e Mccutchan, 2005; Mccutchan, 2006; Requejo, 2006).

O grupo M apresenta uma estrutura filogenética bastante peculiar. A maioria das seqüências referentes a esse grupo pertence a um número limitado de clados equidistantes, o que permitiu a sua classificação em subtipos. Para um isolado pertencer a determinado subtipo, ele deve ser semelhante, ao longo de todo o seu genoma, a um determinado subtipo e a mais nenhum subtipo existente. Baseando-se nesse princípio, o grupo M foi dividido em nove subtipos: A-D, F-H, J e K (Figura 4). Mais recentemente, os subtipos A e F foram divididos em subclados dando origem aos sub-subtipos A, A2, A3, A4, F1 e F2. A distância do pareamento dos nucleotídeos entre os subtipos de HIV-1 é de 20% para a região do envelope, 16% para *gag* e 13% para a região da polimerase. A divergência genética entre os sub-subtipos é de 11-16% para o gene do envelope e de 7-12% para *gag* (Robertson, Anderson *et al.*, 1999; Turner e Summers, 1999; Peeters, 2000; Morgado, Guimaraes *et al.*, 2002; Meloni, Kim *et al.*, 2004; Takeb, Kusagawa *et al.*, 2004; Kijak e Mccutchan, 2005; Mccutchan, 2006; Requejo, 2006; Vidal, Mulanga *et al.*, 2006).

Eventos de recombinação são responsáveis pela formação de genomas mosaicos de HIV-1, compostos por diferentes subtipos, em regiões genômicas distintas. Quando ocorre uma co-infecção por HIV-1 de linhagens distintas, pode haver a formação de vírions cujo genoma é composto por uma fita de RNA de cada linhagem. Ao infectar uma nova célula, o genoma desses vírions pode sofrer recombinação durante a transcrição reversa, por meio da mudança de substrato pela transcriptase reversa, gerando um genoma mosaico. Alguns desses, designados como formas recombinantes circulantes (CRF), foram obtidos de, pelo menos, três isolados epidemiologicamente não relacionados, com os mesmos pontos de passagem entre os subtipos, refletindo a ancestralidade comum dos mesmos eventos de recombinação (Peeters, 2000; Thomson, Perez-Alvarez *et al.*, 2002; Takeb, Kusagawa *et al.*, 2004; Requejo, 2006).

Para definir-se uma nova CRF, é necessário o seqüenciamento do genoma completo das três amostras ou, no mínimo, de duas amostras somadas a uma seqüenciada parcialmente. De acordo com a proposta de nomenclatura para HIV-1, sugerida por Robertson *et al.* (1999), cada CRF deve ser designada por um número de identificação, que segue a ordem cronológica da primeira descrição, e por letras indicando os subtipos envolvidos na recombinação. Se o genoma for formado por seqüências originadas a partir de três subtipos ou mais, as letras podem ser substituídas pela sigla cpx (*complex*) (Robertson, Anderson *et al.*, 1999; Peeters, 2000). Segundo o *Los Alamos HIV Sequence Database* (Los Alamos, 2009b), atualmente, já foram reconhecidas 45 CRF.



**Figura 4. Filogenia dos grupos M, N e O do HIV-1.** O grupo M é formado por clados equidistantes, o que possibilitou sua classificação em subtipos e sub-subtipos. (adaptado de (Thomson, Perez-Alvarez *et al.*, 2002).

### 1.5. Origem do HIV

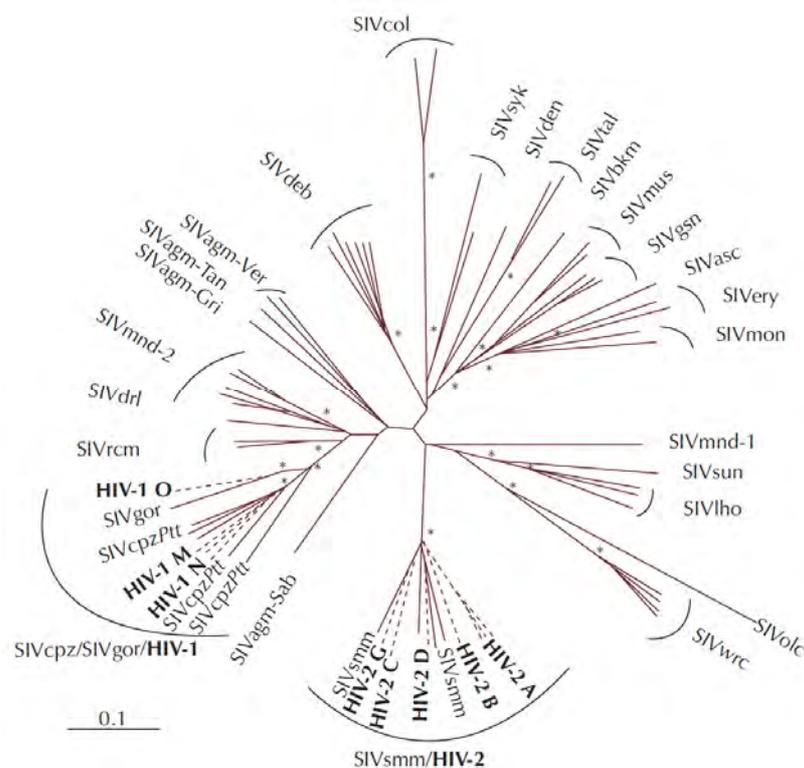
Logo após a identificação do HIV como agente etiológico da AIDS em 1983, foi isolado o primeiro vírus da imunodeficiência em símios (*simian Immunodeficiency virus* – SIV), SIVmac, de macacos rhesus (*Macaca mulatta*) no Centro Regional de Pesquisa de Primatas da Nova Inglaterra (*New England Regional Primate Research Center*). Dez anos depois, macacos da espécie *Macaca arctoides*, do Centro Nacional de Pesquisa de Primatas da Califórnia (*California National Primate Research Center*), desenvolveram uma doença caracterizada por imunossupressão e desenvolvimento de doenças oportunistas, causada por SIVstm. Em ambos os casos, os macacos tiveram contato com sooty mangabeys (*Cercocebus torquatus*) infectados por SIVsmm, no Centro Nacional de Pesquisa de Primatas da Califórnia. Estudos filogenéticos mostraram que as três espécies, SIVmac, SIVstm e SIVsmm, estão estreitamente relacionadas. Baseando-se nessa evidência, passou a se sugerir que SIVsmm tenha dado origem aos SIV em macacos e que o SIVmac, por ter sido relatado em macacos rhesus que apresentam grandes similaridades com humanos, tenha sido o ancestral mais próximo do HIV (Van Heuverswyn e Peeters, 2007).

Análises de distância filogenética revelaram que o HIV-1 e o HIV-2 estão mais próximos a lentivirus que infectam outros primatas do que um do outro. O HIV-1 está fortemente relacionado ao SIVcpz, enquanto o HIV-2 apresenta maior semelhança ao SIVsmm. Além disso, similaridades na organização genômica (presença do gene acessório *vpx*), a coincidência geográfica do epicentro da epidemia HIV-2 e do habitat natural de sooty mangabeys no Oeste da África e o fato dessa espécie de macaco ser frequentemente tratada como animais domésticos e fonte de alimento na região, permitiu a identificação do SIVsmm como ancestral direto do HIV-2 (Holmes, 2001; Van Heuverswyn e Peeters, 2007). O agrupamento disperso das linhagens de HIV-2 e SIVsmm em árvores filogenéticas sustenta a hipótese de que a transmissão espécie-cruzada de SIVsmm para a população humana tenha ocorrido várias vezes e de forma independente (Santiago, Range *et al.*, 2005) na década de 30 (Wertheim e Worobey, 2009) (Figura 5).

De acordo com análises filogenéticas realizadas com os isolados de SIVcpz disponíveis, foram identificadas duas linhagens de SIVcpz de acordo com a sub-espécie hospedeira de chimpanzé: SIVcpzPtt, que infectam *Pan troglodytes troglodytes* do Centro-Oeste africano, e SIVcpzPts, que infectam *Pan troglodytes schweinfurthii* do Leste africano. Todos os grupos do HIV-1, M, N e O, estão mais relacionados aos SIVcpzPtt que aos SIVcpzPts, de modo que esses últimos formam

um clado separado. Assim, elimina-se a possibilidade de o SIVcpzP<sub>ts</sub> ter dado origem ao HIV-1. No entanto, observa-se, ainda, que isolados de SIVcpzP<sub>tt</sub> estão muito mais intimamente relacionados aos grupos M e N que ao O (Holmes, 2001; Paraskevis, Lemey *et al.*, 2003; Van Heuverswyn e Peeters, 2007). Apenas recentemente, com o isolamento de SIV (SIVgor) em gorilas (*Gorilla gorilla*) no Oeste da República de Camarões, África, pode-se provar que, além de SIVgor formarem um grupo monofilético com HIV-1/SIVcpzP<sub>tt</sub>, ele está filogeneticamente mais próximo ao grupo O do HIV-1. Concluiu-se, ainda, que o SIVcpz deu origem aos SIV que circulam atualmente em chimpanzés, gorilas e humanos (Van Heuverswyn, Li *et al.*, 2006) (Figura 5).

A origem do HIV-1 ocorreu no Centro Oeste Africano, na República de Camarões. Além disso, os grupos M e N surgiram de linhagens de SIVcpz geograficamente distintas, tendo o primeiro originado no Sudeste e o segundo no Sul da República de Camarões (Keele, Van Heuverswyn *et al.*, 2006), em 1930 (Korber, Muldoon *et al.*, 2000) e 1963 (Wertheim e Worobey, 2009) respectivamente. O grupo O, por sua vez, originou-se a partir de linhagens de SIVgor no Sudeste da República de Camarões (Van Heuverswyn, Li *et al.*, 2006), em 1920 (Lemey, Pybus *et al.*, 2004).



**Figura 5.** Árvore filogenética representando a relação entre os diferentes vírus da imunodeficiência em símios (SIV) e linhagens do HIV. A análise filogenética foi baseada em seqüências parciais de *pol*, utilizando-se o método *neighbor-joining*. Os asteriscos (\*) indicam valores de *bootstrap* acima de 85% (adaptado de (Van Heuverswyn e Peeters, 2007)).

## 1.6. Distribuição Geográfica do HIV

As formas virais mais prevalentes na epidemia global são os subtipos A e C, seguidos pelo B e pelos recombinantes CRF01\_AE e CRF02\_AG. A maior diversidade genética de HIV-1 é encontrada na África Central. Nesse continente, embora haja uma predominância dos subtipos A e C e do recombinante CRF02\_AG, já foi relatada a ocorrência de todos os grupos e subtipos de HIV-1. Na Ásia, observa-se uma maior prevalência dos subtipos B, C e dos recombinantes CRF01\_AE, CRF15\_01B e CRF8\_BC. Na Europa Ocidental, no início da epidemia HIV/AIDS, o subtipo B era o mais prevalente. No entanto, o contato contínuo com o continente africano por meio de viagens e imigração resultou em um aumento da variabilidade das linhagens de HIV circulantes na Europa, sendo marcante a presença dos subtipos A, C e G e das formas recombinantes CRF02\_AG e CRF\_03AB. Na América do Norte e Austrália, o subtipo B é o predominante (Peeters, 2000; Osmanov, Pattou *et al.*, 2002; Takeb, Kusagawa *et al.*, 2004; Kijak e Mccutchan, 2005; Hemelaar, Gouws *et al.*, 2006; Requejo, 2006).

Na América do Sul, observa-se uma predominância do subtipo B, seguido pelo F na Argentina, Bolívia, Uruguai, Peru, Paraguai e Venezuela (Hierholzer, Montano *et al.*, 2002). No entanto, estudos epidemiológicos conduzidos na Argentina, Uruguai e Bolívia, realizaram o seqüenciamento do genoma completo do HIV-1 e revelaram que a maior parte dos vírus classificados anteriormente como F são, na verdade, recombinantes únicos B/F e CRF\_12BF (Carr, Avila *et al.*, 2001; Thomson, Delgado *et al.*, 2002). Essas formas recombinantes também foram descritas na Venezuela (Castro, Echeverria *et al.*, 2003). A presença do subtipo C foi relatada na Argentina, Paraguai, Uruguai e Venezuela (Carrion, Eyzaguirre *et al.*, 2004; Castro, Moreno *et al.*, 2005) além de casos isolados de A no Chile, Peru e Guiana Francesa (Morgado, Guimaraes *et al.*, 2002). Dados da Colômbia e Equador só revelaram a presença do subtipo B (Morgado, Guimaraes *et al.*, 2002) (Figura 6). Devido escassez de trabalhos abordando a epidemia do HIV-1 nesses dois últimos países, a identificação apenas do subtipo B pode ser decorrente da subnotificação de subtipos prevalentes na região, sendo provável a circulação de outros subtipos e formas recombinantes.



**Figura 6. Distribuição geográfica dos subtipos e formas recombinantes circulantes (CRF) do HIV-1 no mundo.** North and Central America, América do Norte e Central; South America, América do Sul; Western Europe, Europa Ocidental; Eastern Europe, Europa Oriental; Western Africa, África Ocidental; East Africa, África Oriental; Central Africa, África Central; South África, Sul da África; South Asia, Sul da Ásia; Southeast Asia, Sudeste da Ásia (adaptado de Thomson, Perez-Alvarez *et al.*, 2002).

A epidemia HIV/AIDS no Brasil é complexa quanto à distribuição e prevalência dos diferentes subtipos, se comparada aos outros países da América Latina. O subtipo B do HIV-1 tem sido descrito como o mais prevalente no Brasil, seguido pelo F e C, respectivamente. Além disso, foram relatados alguns casos isolados de A, D e genomas mosaicos em potencial, envolvendo recombinação ou infecção dupla entre B/F, B/C e F/D. Em adição a diversidade inter-subtipo, diferenças genéticas e antigênicas também foram descritas entre linhagens do subtipo B circulantes no Brasil, com a identificação de um variante denominado B''. Esse difere do subtipo B clássico pela presença do motivo GWGR no topo do *loop* V3 de gp120 do envelope no lugar de GPGR. Em algumas áreas do Brasil, o variante B'' mostrou-se altamente prevalente, correspondendo a 57% dos subtipos B detectados em Ribeirão Preto (SP) e 37% dos do Rio de Janeiro (RJ) (Morgado, Guimaraes *et al.*, 2002) (Figura 7).

Em um estudo conduzido no Norte do país, foi descrita a presença do subtipo B em 38,7% das amostras analisadas, seguida por uma alta porcentagem de recombinantes B/F (35,5%) e do subtipo F (16%) (Vicente, Otsuki *et al.*, 2000). Já no Nordeste, foi registrada uma prevalência de, aproximadamente, 80% de subtipo B,

com casos esporádicos dos subtipos F e C e recombinantes B/F (Couto-Fernandez, Morgado *et al.*, 1999; Gadelha, Shindo *et al.*, 2003; Monteiro, Alcantara *et al.*, 2009).

No Centro-Oeste, há uma alta prevalência de subtipo B (aproximadamente 90%), seguido por recombinantes B/F (cerca de 5%) (Cerqueira, Amorim *et al.*, 2004; Veras, Veras *et al.*, 2007). Mais recentemente, foi descrita a presença do subtipo C (1,5%) em Goiás e de formas recombinantes (14,5%) envolvendo os subtipos B, C, D e F (Stefani, Pereira *et al.*, 2007).

Estudos realizados no Sudeste e Sul do Brasil mostraram uma grande proporção de infecções pelo subtipo C do HIV-1 nos estados do Paraná (30%) (Ferreira, Thomaz *et al.*, 2008), Santa Catarina (49-64%) (Brigido, Nunes *et al.*, 2007; Locateli, Stoco *et al.*, 2007) e Rio Grande do Sul (27-40%) (Soares, Santos *et al.*, 2003; Brigido, Nunes *et al.*, 2007). A prevalência deste subtipo também aumentou no Espírito Santo (Cabral, Cunha *et al.*, 2006), Rio de Janeiro (Dumans, Soares *et al.*, 2002; Pires, Soares *et al.*, 2004) e São Paulo (De Oliveira, Almeida *et al.*, 2008; De Sa-Filho, Soares Mda *et al.*, 2008; De Sa-Filho, Ambar *et al.*, 2009), sugerindo uma dispersão do subtipo C em direção ao Norte do país. Nesses estados, também foi descrita a presença dos subtipos B, B'' e F1, e dos recombinantes B/F, B/D e B/C .



**Figura 7. Distribuição geográfica dos subtipos e formas recombinantes do HIV-1 no Brasil** (modificado de (Morgado, Guimaraes *et al.*, 2002).

## 1.7. Anti-retrovirais e Resistência

No Brasil, em 13 de novembro de 1996, foi sancionada a lei nº 9.313 que disponibiliza o acesso gratuito aos anti-retrovirais de alta potência (terapia anti-retroviral altamente ativa - HAART) para pacientes infectados por HIV (Brasil, 1996). O principal objetivo da terapia anti-retroviral é manter a replicação do HIV-1 em níveis baixos (<50 cópias/mL), retardando a progressão da imunodeficiência e/ou restaurando, tanto quanto possível, a imunidade. Assim garante-se uma melhor qualidade de vida ao paciente e redução da transmissão do vírus assim como a morbidade e mortalidade associada ao HIV (Brasil, 2008b; Department of Health and Human Services, 2009).

De acordo com o Departamento de Saúde e Serviços Humanos dos Estados Unidos da América, o tratamento anti-retroviral deve ser iniciado em todos os pacientes infectados pelo HIV que se enquadrem nas seguintes categorias (Department of Health and Human Services, 2009):

- sintomáticos;
- gestantes;
- apresenta nefropatia associada a HIV;
- co-infectados pelo vírus da hepatite B (HBV), quando o tratamento para HBV também for indicado; e
- apresenta contagem de células T CD4<sup>+</sup> inferior a 350 células/mm<sup>3</sup>.

No Brasil, de acordo com o Ministério da Saúde, o tratamento anti-retroviral é recomendado para todos os pacientes infectados pelo HIV que sejam sintomáticos, independentemente da contagem de linfócitos T CD4<sup>+</sup>, e para aqueles assintomáticos com contagem de linfócitos T CD4<sup>+</sup> abaixo de 200 células/mm<sup>3</sup>. No caso de pacientes com contagem de linfócitos T CD4<sup>+</sup> entre 200 e 350 células/mm<sup>3</sup>, o início da terapia anti-retroviral pode ser considerado de acordo com a evolução dos parâmetros imunológicos (contagem de linfócitos T CD4<sup>+</sup>) e virológicos (carga viral acima de 100.000 cópias/mL), levando-se, também, em consideração a motivação e capacidade de adesão ao tratamento por parte do paciente (Brasil, 2008b).

Atualmente, existem 25 agentes anti-retrovirais formalmente aprovados para o tratamento da infecção por HIV, pertencentes a seis classes de medicamentos: inibidor de protease (PI), inibidor da transcriptase reversa análogo de nucleosídeos ou nucleotídeos (NRTI) e não análogo de nucleosídeos (NNRTI), inibidor de fusão (FI), inibidor de co-receptor (CRI) e inibidor de integrase (II) (Cane, 2009; De Clercq, 2009; Adams, Patel *et al.*, 2010; Schafer e Squires, 2010).

Os PI exibem uma alta afinidade pelo sítio catalítico da protease, inibindo a atividade enzimática proteolítica, o que resulta na geração de partículas imaturas não-infectivas e, conseqüentemente, em uma dramática supressão da replicação viral. Até o momento, 10 inibidores de protease foram aprovados pelo órgão do governo americano de regulamentação de drogas e alimentos (*food and drug administration – FDA*): atazanavir (ATV), amprenavir (APV), darunavir (DRV), fosamprenavir (FPV) indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), ritonavir (RTV), saquinavir (SQV) e tipranavir (TPV) (Gallant, 2002; Martinez-Cajas e Wainberg, 2007; De Clercq, 2009).

Os NRTI, mimetizam a estrutura dos nucleosídeos ou nucleotídeos naturais de modo que, ao ser adicionado na fita de DNA pela transcriptase reversa, promove o término precoce da transcrição, permitindo que enzimas celulares destruam as partículas virais. Os análogos de nucleosídeos são, classificados em dois subgrupos: análogos de timidina e não-análogos de timidina. O primeiro atua, preferencialmente, contra o HIV em células CD4<sup>+</sup> ativadas e é representado pelo zidovudina (AZT) e estavudina (d4T). O segundo grupo é composto pelo didanosina (ddI), zalcitabina (ddC), lamivudina (3TC), emtricitabina (FTC) e abacavir (ABC) e apresenta atividade equivalente em células CD4<sup>+</sup> ativadas ou não. Vale ressaltar que o ddC não é mais comercializado. O único análogo de nucleotídeo aprovado pelo FDA é o tenofovir (TDF). Ele atua de forma semelhante ao análogo de nucleosídeo. No entanto, esse inibidor já é monofosforilado, facilitando sua ativação intracelular por meio de fosforilação. O TDF é ativo em uma série de células, incluindo linfócitos e macrófagos não ativados (Gallant, 2002; De Clercq, 2009).

Os NNRTI inibem a replicação do HIV por se ligarem, não competitivamente, à transcriptase reversa, em uma região hidrofóbica próxima ao sítio ativo da polimerase, causando uma mudança alostérica, que provoca a ligação não produtiva dos desoxinucleotídeos. Esse grupo de antiretrovirais é composto pelo nevirapina (NVP), delavirdina (DLV), efavirenz (EFV) e etravirina (ETR) (Gallant, 2002; De Clercq, 2009; Adams, Patel *et al.*, 2010).

Os FI se ligam a gp41 no momento em que ocorre sua mudança estrutural necessária para haver a fusão do envelope e, conseqüentemente, a entrada do vírus na célula. O único FI atualmente aprovado é o enfuvirtide (T20 ou ENF) o qual mimetiza a porção HR2 de gp41 (aminoácidos 127 a 162) impedido sua ligação a porção HR1, necessária para que ocorra a fusão entre a membrana celular e o envelope viral (Este e Telenti, 2007; Shafer e Schapiro, 2008).

No momento, apenas um CRI foi licenciado para uso clínico, o maraviroc. Esse anti-retroviral é um antagonista ao CCR5 que impede a ligação da glicoproteína viral gp120 ao co-receptor CCR5, impossibilitando a entrada do vírus na célula. Uma vez que esse medicamento só impede eficientemente a entrada de linhagens virais com tropismo ao co-receptor CCR5 os quais, por sua vez, estão em maior abundância na fase inicial da infecção, esse medicamento apresenta maior utilidade quando administrado na fase aguda da doença (Este e Telenti, 2007; Perno, Moyle *et al.*, 2008; Shafer e Schapiro, 2008; Bhattacharya e Osman, 2009).

Por sua vez, o raltegravir é o único II aprovado pelo FDA. Esse agente se liga ao sítio catalítico da integrase, interagindo com os co-fatores de cátions bivalente. Desse modo, a ligação do complexo pré-integração é inibida, impedindo a transferência da fita de DNA viral ao DNA celular (Shafer e Schapiro, 2008; Schafer e Squires, 2010).

Um importante obstáculo encontrado na terapia da AIDS tem sido o rápido aparecimento de cepas virais resistentes. A resistência do HIV-1 pode ser definida como qualquer mudança que melhore a replicação viral na presença de um inibidor. Essa resistência é sempre relativa ao vírus tipo-selvagem e representa um fenótipo alterado resultante de uma mudança no genótipo do vírus. Como o HIV-1 tem altas taxas de mutação e de replicação, variantes resistentes às drogas podem ser gerados aleatoriamente, estando presentes em pequenas concentrações mesmo antes da terapia ser iniciada (Dumans, Soares *et al.*, 2002; Beerenwinkel, Daumer *et al.*, 2005).

A resistência aos PI está associada a um padrão complexo de mutações pontuais e o aparecimento de resistências fenotípicas significativas requer uma combinação de mutações, que emergem de modo seqüencial. As mutações que conferem resistência aos inibidores de protease podem ser classificadas como primárias ou secundárias. As primárias são as primeiras a serem selecionadas na presença da droga e conferem mudanças fenotípicas que prejudicam a ação do anti-retroviral. Normalmente ocorrem no sítio de ligação da droga. Já as secundárias, emergem tardiamente, em relação às primárias. Elas, por si só, não apresentam um efeito significativo no fenótipo e podem, quando em conjunto com mutações primárias, melhorar a capacidade de replicação do vírus (Gallant, 2002; Martinez-Cajas e Wainberg, 2007; Johnson, Brun-Vezinet *et al.*, 2009).

A resistência aos NRTI é desencadeada por mutações no gene da RT que alteram a discriminação entre os NRTI e os dNTP ou aumentam a atividade de fosforilação da RT que, na presença de ATP, permite a remoção de nucleotídeos na

terminação 3' da cadeia de DNA. A emergência de diferentes mutação no sítio de ligação ao dNTP ou em proximidade a ele está relacionada à diminuição de susceptibilidade aos NRTI (Martinez-Picado e Martinez, 2008). Já no caso dos NNRTI, mutações pontuais no sítio alostérico de ligação da droga à enzima é responsável pelo surgimento de linhagens resistentes a essa classe de medicamentos. Uma vez que o sítio de ligação é semelhante para os diferentes NNRTI, é muito comum haver resistência cruzada de modo que uma única mutação pode provocar resistência à toda a classe (Adams, Patel *et al.*, 2010).

A redução de susceptibilidade ao inibidor de fusão T20 é ocasionada principalmente por mutações que ocorrem nos códons 36 a 45 de gp41, local onde se liga o anti-retroviral. Uma única mutação nesse sítio pode ocasionar uma diminuição na susceptibilidade à droga em até 10 vezes, enquanto a presença de duas mutações pode diminuir em mais de 100 vezes. Várias mutações acessórias em HR2 pode auxiliar no desenvolvimento de resistência quando na presença de mutações nos aminoácidos 36 a 45 (Este e Telenti, 2007; Shafer e Schapiro, 2008).

Devido à sua especificidade a CCR5, a resistência ao maraviroc pode ser resultado do aparecimento de linhagens com tropismo ao CXCR4 (X4) ou com tropismo duplo (R5/X4) e pela presença de mutações em gp120 permitindo a ligação do vírus ao co-receptor mesmo na presença do inibidor. As mudanças de aminoácidos responsáveis pela resistência podem estar localizadas no *loop* V3, fora do *loop* V3 ou resultar de interações sinérgicas entre substituições no *loop* V3 e outras partes de gp120. Essas mudanças podem ainda aparecer como polimorfismos, inserções ou deleções (Shafer e Schapiro, 2008; Cane, 2009; Marcelin, Ceccherini-Silberstein *et al.*, 2009).

Por fim, a resistência ao II raltegravir está associada a presença de mutações localizadas principalmente no sítio de ligação do anti-retroviral à enzima. Três vias principais de resistência já foram identificadas e envolvem as substituições Y143R/C, Q148K/R/H e N155H. Mutações secundárias também podem diminuir a susceptibilidade a esse medicamento (Shafer e Schapiro, 2008; Cane, 2009; Marcelin, Ceccherini-Silberstein *et al.*, 2009; Schafer e Squires, 2010).

## **1.8. Resposta Imune e Desenvolvimento de Vacinas**

Tanto a resistência à infecção por HIV quanto a progressão da doença dependem do sistema imune do hospedeiro, que consiste em duas principais vias de defesa: inata e adquirida. A imunidade inata, mediada por fagócitos, células

dendríticas, células *natural killer* (NK) e sistema complemento, atua rapidamente e desempenha um importante papel no controle inicial da infecção. Já a imunidade adquirida consiste na resposta imune humoral, mediada por linfócitos B, e celular, mediada por linfócitos T (Gandhi e Walker, 2002; Levy, Scott *et al.*, 2003).

A infecção por HIV inicia-se em um foco de células T CD4<sup>+</sup>. A replicação viral nesse foco é, então, sustentada pela resposta imune inata que desencadeia o recrutamento de mais células T susceptíveis ao local da infecção (Mcmichael, Borrow *et al.*, 2010). A principal função do sistema imune inato envolve a ativação de uma via de sinalização intracelular por meio de receptores localizados em uma variedade de células, sendo a maioria apresentadora de antígenos (células dendríticas, macrófagos). Esses receptores (*tool-like receptors* – TLR), após interagirem com ligantes específicos, iniciam uma cascata que leva à ativação do fator NFκB e indução de citosinas. Desse modo, a resposta imune inata ativa a produção de uma série de fatores celulares e aumenta a resposta de células do próprio sistema imune inato além de recrutar células T CD4<sup>+</sup> para o lugar da infecção. Além disso, também induz a resposta imune adaptativa aumentando a expressão de complexo maior de histocompatibilidade (*major histocompatibility complexes* – MHC) em células alvo em potencial e aumenta a resposta dos linfócitos aos patógenos (Gandhi e Walker, 2002; Levy, Scott *et al.*, 2003).

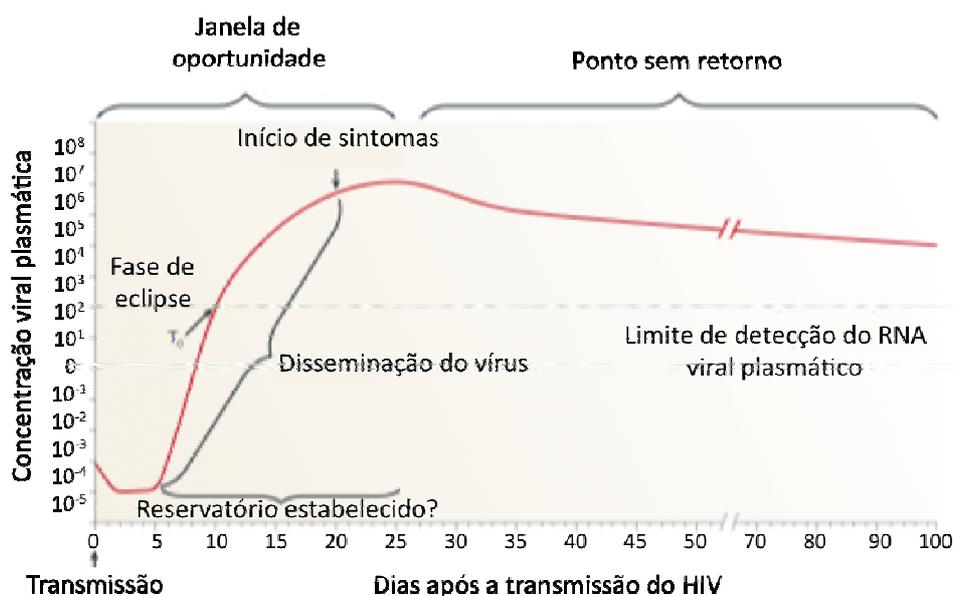
A fase de eclipse, a qual ocorre durante os 5-10 primeiros dias após a infecção, antes que as partículas virais e/ou células infectadas atinjam os nódulos linfáticos, corresponde ao período em que uma vacina preventiva poderia erradicar o vírus do organismo. O fato de a infecção ser resultado da transmissão de um variante viral favorece o ataque imune durante a fase de eclipse, de modo que uma vacina poderá controlar a viremia e evitar danos aos componentes e ativação do sistema imune. Ao atingirem os nódulos linfáticos, o vírus passa a infectar células T CD4<sup>+</sup> CCR5<sup>+</sup> ativadas, estabelecendo, portanto, reservatórios de vírus latente (Mcmichael, Borrow *et al.*, 2010). Esse processo é aumentado pelas células dendríticas (DC) que possuem o receptor DC-SIGN. Essas células são altamente prevalentes em mucosas e se ligam ao vírus por meio do receptor DC-SIGN, auxiliando no transporte do vírus até órgãos linfáticos secundários onde o vírus é transferido para células T (Geijtenbeek, Kwon *et al.*, 2000). O mesmo receptor foi encontrado em células B no sangue periférico e órgãos linfáticos, estando também envolvido na trans-infecção de células T (Rappocciolo, Piazza *et al.*, 2006). Além do receptor DC-SIGN, a presença de CD21

em células B também está envolvida na transferência de vírus para células T (Moir, Malaspina *et al.*, 2000).

O vírus passa a se replicar rapidamente e se espalhar pelo organismo, atingindo outros tecidos linfóides, especialmente aquele associado ao intestino (GALT), onde células T de memória  $CD4^+ CCR5^+$  ativadas estão presente em grande quantidade. Nesse período, a carga viral plasmática aumenta exponencialmente até atingir o ápice em 14-21 dias após a infecção. As primeiras respostas das células T  $CD8^+$  ocorrem nesse período. No entanto, a redução de células T  $CD4^+$  compromete a atividade dos linfócitos T  $CD8^+$ . Uma vacina deveria, portanto, estimular apropriadamente a resposta de linfócitos T  $CD4^+$  para que os mesmos possam melhor sustentar a ação de células T  $CD8^+$  a vários epitopos conservados, favorecendo o reconhecimento de vários variantes virais (Watkins, 2008; Mcmichael, Borrow *et al.*, 2010).

A carga viral passa, então, a diminuir lentamente até atingir um nível mais estável após 12 a 20 semanas. Durante essa fase, ocorre a diversificação viral e vários mutantes são selecionados sob a pressão da resposta imune adaptativa, a qual é detectada logo após a carga viral atingir o ápice (Mcmichael, Borrow *et al.*, 2010).

A Figura 8 ilustra os principais eventos que ocorrem durante a fase aguda da infecção por HIV-1, indicando o momento em que a aplicação de uma vacina preventiva poderia eliminar o vírus do organismo. É importante ressaltar que o modelo apresentado refere-se a aplicação de uma vacina ainda hipotética e que, portanto, o limiar entre janela de oportunidade e ponto sem retorno ainda não pôde ser de fato comprovado. Além disso, excelentes resultados terapêuticos já foram alcançados durante esse período da infecção.



**Figura 8. Principais eventos durante a fase aguda da infecção por HIV-1.** A fase de eclipse corresponde ao período inicial da infecção até que ocorra a disseminação viral sistêmica. O momento que o vírus passa a ser detectável no sangue é considerado o  $T_0$ , após o qual, a carga viral plasmática passa a crescer exponencialmente atingindo o pico em 21 a 28 dias após a infecção. Durante o pico de viremia, o indivíduo pode tornar-se sintomático e reservatórios de vírus latente são estabelecidos. A janela de oportunidade corresponde ao período crucial durante o qual a aplicação de uma vacina anti-HIV pode controlar a infecção, prevenindo a depleção de células  $T\ CD4^+$  (adaptado de (Mcmichael, Borrow *et al.*, 2010).

Após um a três meses de infecção, os pacientes soropositivos apresentam uma alta concentração de anticorpos direcionados para diversas proteínas virais. Os primeiros anticorpos contra HIV detectados reconheciam proteínas estruturais como da matriz (p17) e capsídeo (p24). Em seguida, foram descritos anticorpos que reconheciam epítopos em proteínas do envelope e da polimerase. Os anticorpos neutralizantes são aqueles que neutralizam a partícula viral antes que esta entre na célula. Teoricamente, esses anticorpos podem exercer uma atividade antiviral e, normalmente, agem contra a região hipervariável da glicoproteína do envelope, gp120, no sítio de ligação a CD4 e receptores de quimiocinas (Gandhi e Walker, 2002; Goepfert, 2003). A sobreposição de *loops* altamente variáveis (V1 e V2 de gp120) dificulta o acesso a sítios de gp120 críticos para a ligação ao receptor celular, e a glicosilação do envelope pode mascarar epítopos importantes dificultando a neutralização das diferentes linhagens de HIV (Montefiori e Mascola, 2009).

A resposta imune adaptativa celular inclui a atividade de linfócitos T auxiliares e linfócitos T citotóxicos (CTL). Para que ocorra, é necessário que as células T apresentem o antígeno ligado a proteínas celulares conhecidas como MHC ou antígenos de leucócitos humanos (*human leukocyte antigens* – HLA). A grande maioria dos CTL são  $CD8^+$  e, portanto, reconhecem peptídeos do patógeno composto

por 9 a 11 aminoácidos ligados a moléculas de MHC de classe I. Essa proteína está presente em todos os tipos de células humanas. Já os linfócitos T CD4<sup>+</sup> reconhecem peptídeos com 11 a 20 aminoácidos associados a MHC de classe II. Essas, por sua vez, são encontradas apenas em células apresentadoras de antígenos como macrófagos, células dendríticas e linfócitos. Essas células são capazes de fagocitar e processar proteínas exógenas em pequenos peptídeos que são, subsequentemente, apresentados por MHC de classe II e reconhecidos por células T CD4<sup>+</sup>. Sendo assim, via de regra, as células T CD8<sup>+</sup> respondem a antígenos sintetizados no interior celular e que, portanto, requerem a infecção da célula; enquanto que os linfócitos T CD4<sup>+</sup> respondem a antígenos encontrados no meio extracelular. Como as infecções virais são compostas por antígenos endógenos e exógenos, ambas respostas celulares são ativadas (Gandhi e Walker, 2002; Goepfert, 2003).

Os CTL podem controlar a infecção por HIV por meio de, pelo menos, dois mecanismos. Primeiro, CTL específicos para vírus podem lisar células infectadas por HIV ao reconhecer peptídeos virais na superfície celular conjugadas a moléculas MHC de classe I. Uma outra possibilidade é a inibição da replicação do HIV por meio da secreção de quimocinas  $\beta$ , como MIP-1 $\alpha$ , MIP-1 $\beta$  e RANTES, por células CD8<sup>+</sup>. Essas moléculas se ligam aos co-receptores do HIV-1, bloqueando a entrada do vírus, e, também, promovem a atividade citolítica de CTL, potencializando o controle do HIV-1 por células T CD8<sup>+</sup> (Gandhi e Walker, 2002; Goepfert, 2003).

As células T CD4<sup>+</sup> parecem desempenhar papel fundamental na manutenção de uma resposta imune efetiva. Pacientes com os maiores índices de atividade de células T CD4<sup>+</sup> apresentam menores cargas virais e vice-versa. Nesses pacientes a resposta de linfócitos T auxiliares está associada à resposta de CTL. Assim, as células T CD4<sup>+</sup> devem coordenar a resposta imune contra HIV por, no mínimo, manter efetiva a atividade de CTL (Gandhi e Walker, 2002; Goepfert, 2003).

Os tratamentos disponíveis atualmente não curam, mas apenas desaceleram a progressão da doença. A maioria dos medicamentos necessita de acompanhamento médico periódico, é de difícil administração, causa vários efeitos colaterais e o custo é extremamente elevado, sendo inacessível a uma grande parte dos indivíduos infectados, principalmente nos países em desenvolvimento. Dessa forma, uma das melhores soluções para controle da AIDS, em longo prazo, é o desenvolvimento de uma vacina segura, efetiva e acessível. A vacina preventiva contra HIV ideal deve ser efetiva independentemente do estado de saúde e nutricional da população alvo, deve proteger os indivíduos contra todos os subtipos e qualquer via de transmissão, ser fácil

de transportar e administrar, proporcionar longo período de proteção e ser de baixo custo (Iavi, 2008).

Atualmente, mais de 20 vacinas candidatas estão sendo testadas em triagens clínicas de pequena escala. A primeira triagem de larga escala foi concluída em 2003 com a candidata VaxGen's AIDSVAX, que foi testada na América do Norte e na Tailândia. Embora não considerada efetiva, proporcionou grandes avanços no estudo de vacinas contra HIV. Após 25 anos de pesquisa para a produção de uma vacina efetiva contra o HIV, esforços estão sendo direcionados para novas estratégias que permitam acelerar o desenvolvimento de uma vacina. No momento, somente a vacina baseada na proteína gp120 utilizando canarypox como vetor (Sanofi Pasteur) está sendo testada em triagem de larga escala (fase III) na Tailândia. Existe uma tendência de triagens de larga escala serem substituídas por várias menores, preliminares, com, apenas aquelas que indicarem evidências e eficácia sendo testadas em maior escala (Iavi, 2008).

Apesar do avanço na pesquisa para o desenvolvimento de uma vacina efetiva contra AIDS, existe uma série de obstáculos que dificultam sua concretização. O primeiro é que a vacina não pode consistir em vírus atenuado, ou mesmo inativo, devido ao perigo inerente de os mesmos causarem AIDS, já que não se pode assegurar que todas as partículas estarão, de fato, inativas. Desse modo, a pesquisa fica restrita ao uso de proteínas ou peptídeos virais ou a partes do genoma, associados ou não a um vetor. Um segundo obstáculo é que não existe um modelo animal de pequeno porte para os experimentos, limitando-os ao uso de macacos infectados por SIV, que, nem sempre, são de fácil acesso. Desse modo, os resultados obtidos com SIV são extrapolados para HIV. O terceiro problema aparente é que não se sabe ao certo qual resposta imune irá promover melhor proteção. A maioria dos protótipos tem investido nas respostas humoral e celular. Poucos estudos investem na resposta imune inata, da mucosa ou na utilização de quimiocinas que inibam o HIV como um suplemento para desencadear uma resposta imune, de fato, eficiente (Gallo, 2005; Girard, Osmanov *et al.*, 2006).

Uma outra dificuldade encontrada é a variabilidade das linhagens de HIV. Até hoje, não se sabe a influência dessa diversidade na ação efetiva de uma vacina. Portanto, até que essa questão seja respondida, devem-se considerar os diferentes variantes de HIV prevalentes no mundo e não restringir a pesquisa ao subtipo B. Embora seja o mais prevalente na América do Norte, América do Sul, Europa, Austrália e Nova Zelândia, o subtipo B representa uma pequena fração das infecções

por HIV no mundo. Os subtipos A e C representam, juntos, dois terços das infecções e, mesmo assim, não vêm sendo utilizados na fabricação de vacinas (Gallo, 2005; Girard, Osmanov *et al.*, 2006; Iavi, 2008).

No entanto, talvez o obstáculo mais importante seja o fato de o HIV ser um retrovírus e, portanto, integrar seu genoma ao DNA celular. Dessa forma, estabelece-se uma infecção crônica, caso a intervenção não ocorra no momento da exposição inicial. Além disso, o HIV afeta não só as células infectadas como também as vizinhas. Esse efeito é mediado pela proteína Tat, que é liberada pelas células infectadas. Cientes disso, alguns pesquisadores focaram seus estudos na indução de anticorpos neutralizantes contra a glicoproteína do envelope gp120, tentando bloquear a entrada do vírus na célula. O problema foi que a proteção só foi efetiva para a linhagem utilizada na produção da proteína ou linhagens extremamente próximas (Gallo, 2005; Girard, Osmanov *et al.*, 2006).

Alguns desses problemas podem ser solucionados mais facilmente. Por exemplo, a idéia da produção de uma vacina de subunidade é bem aceita e essa estratégia já mostrou ser eficiente contra outros vírus. Os modelos de macacos expostos a SIV podem ser mais acessíveis. Isso depende da parte burocrática junto aos centros de primatologia. Embora haja uma grande variabilidade entre as linhagens de HIV, existem regiões mais conservadas, inclusive envolvidas na interação de gp120 com CD4 ou CCR5. Ao bloquear a entrada do vírus na célula, a integração do genoma viral ao DNA celular também será evitada (Gallo, 2005).

Paralelamente ao desenvolvimento de vacinas preventivas, esforços têm sido feitos para o desenvolvimento de vacinas terapêuticas. Estas consistem na vacinação de indivíduos imunocompetentes (contagem de CD4 inferior a 350 células/mm<sup>3</sup>), com baixa viremia (carga viral menor que 50 cópias/mL), que estão sob tratamento com anti-retrovirais (HAART), visando estabelecer uma imunidade celular HIV-específica, conter a replicação viral e atrasar o aumento da carga viral após um período sem o uso de anti-retrovirais. Várias triagens clínicas de fase I e II estão sendo realizadas e resultados preliminares indicam que as vacinas são bem toleradas e conseguem restabelecer a proliferação de linfócitos T, ativando uma resposta celular HIV-específica e suprimindo a replicação viral (Klein, 2003; Lu, Arraes *et al.*, 2004; Levy, Gahery-Segard *et al.*, 2005).

## 1.9. Evolução Molecular

As mutações que ocorrem em um gene e são passadas para os descendentes e que co-existem com o gene original resultam em polimorfismos. Considerando um sítio polimórfico, dois ou mais variantes de um gene podem circular simultaneamente em uma população. Cada um desses variantes é chamado de alelo. A evolução é resultado da mudança na frequência dos alelos. Para vírus de RNA, especificamente, a evolução é reflexo da frequência de um variante na distribuição das quasispecies (Vandamme, 2009).

A taxa com que as mutações são fixadas em uma população é chamada de taxa evolutiva (número de nucleotídeos mudados por sítio por ano), a qual depende da taxa de mutação (número de nucleotídeos mudados por sítio por ciclo de replicação), do tempo que separa duas gerações (tempo de geração) e das forças evolutivas (pressões seletivas positivas e negativas, tamanho populacional e potencial reprodutivo) (Vandamme, 2009).

A árvore filogenética é uma forma elegante de se ilustrar as relações evolutivas entre genes e organismos. Em uma árvore filogenética, os ramos externos (taxa existente) são denominados de unidades taxonômicas operacionais (UTO) e os internos, de unidades taxonômicas hipotéticas (UTH), para enfatizar que são os progenitores hipotéticos. As UTO podem ter se originado de ancestrais diferentes, na geração anterior, ou de um mesmo ancestral. Nesse último caso, trata-se de um evento de coalescência. A data da existência do ancestral comum mais recente das UTO é determinada como tempo de coalescência. Em análises que estimam a origem de uma população por meio da variabilidade de diferentes alelos de um gene em particular, o tempo de coalescência dependerá da extinção dos alelos após especiação (Slatkin e Maddison, 1989; Kingman, 2000; Drummond, Rambaut *et al.*, 2005; Vandamme, 2009).

Para estimar o tempo de coalescência entre genes, alelos ou variantes de quasispecies, normalmente assume-se que a divergência das seqüências aumente ao longo do tempo de forma linear. Essa é a premissa do relógio molecular, o qual considera que todas as linhagens em uma árvore acumulam substituições em uma mesma taxa (taxa evolutiva constante). No entanto, a taxa evolutiva depende de muitos fatores como taxa metabólica na espécie, tempo de geração e pressão evolutiva, de modo que um relógio molecular absoluto não existe. Principalmente para vírus, que possuem altas taxas de replicação; mudanças de hospedeiro, e conseqüentemente de pressão seletiva do ambiente; e que estão constantemente

sofrendo efeito de “gargalo de garrafa”, sempre há diferença nas taxas evolutivas dos ramos de uma árvore. No entanto, testes estatísticos podem fornecer uma idéia do quão distante essas taxas evolutivas estão de uma taxa uniforme e, muitas vezes, a aplicação de relógio molecular é viável. Mais recentemente, o uso do relógio molecular relaxado, o qual permite que a taxa evolutiva varie entre os ramos da árvore, tem se mostrado um método mais acurado para inferir relações filogenéticas e datar o tempo de divergência a partir do ancestral em comum (Drummond, Ho *et al.*, 2006; Vandamme, 2009).

A alta taxa evolutiva de alguns vírus, em especial os de RNA, faz com que os processos epidemiológicos e ecológicos que moldam a diversidade genética ocorram em uma mesma escala cronológica que a fixação de mutações na população viral. Conseqüentemente, filogenias construídas a partir de seqüências de genes virais escolhidas randomicamente contêm informações preciosas quanto a mudanças no tamanho populacional e taxa de crescimento, que podem ser utilizadas para entender o curso de uma epidemia viral ao longo do tempo. Nesse contexto, a filogeografia aparece como uma importante área da biologia evolutiva moderna, promovendo o estudo da biogeografia por meio da comparação de árvores filogenéticas de diferentes populações virais contendo a informação da distribuição geográfica dessas populações. O estudo da filogeografia e evolução viral pode fornecer predições importantes quanto ao aparecimento de novas infecções, a espécie que funciona como reservatório de novas linhagens virais, e o local de origem e dispersão dessas linhagens na população humana (Holmes, 2008).

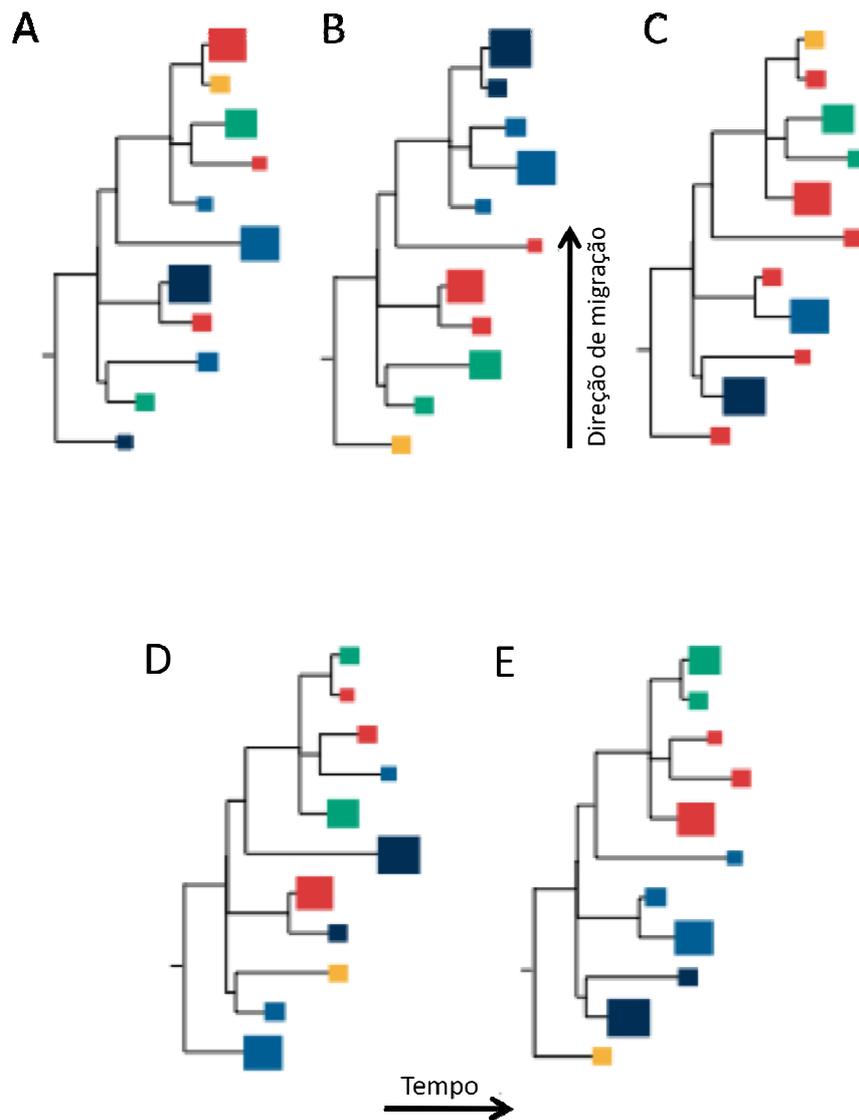
Diferentes padrões filogeográficos são observados em vírus que infectam humanos (Holmes, 2008) (Figura 9):

- Ausência de estrutura espacial: mistura randômica e complexa entre isolados de diferentes regiões geográficas, indicativo de circulação freqüente de linhagens virais entre as localidades.
- Trasmissão do tipo onda: o vírus se dispersa a partir de um local inicial central. Exemplo: Influenza (transmissão no Brasil ocorre na direção Norte-Sul), vírus do sarampo.
- Modelo fonte-dreno (core-satélite): um ou um número limitado de áreas geográficas atuam como fonte, exportando linhagens virais para outras regiões onde o vírus sobrevive por um curto intervalo de tempo e a partir das quais pode gerar outras ondas de transmissão. Exemplo: Influenza (transmissão global).

- Dinâmica gravitacional: a transmissão viral é dirigida por grandes centros populacionais que atuam como atrator gravitacional. Exemplo: Influenza (transmissão nos Estados Unidos da América tende a seguir o movimento das pessoas para e a partir dos locais de trabalho).
- Forte subdivisão espacial: diferentes lugares são caracterizados pela presença de isolados virais filogeneticamente distintos com pouca evidência de fluxo gênico entre esses locais. Exemplo: HIV e vírus da hepatite C (HCV).

Esses padrões de migração refletem a taxa relativa de fluxo gênico viral, o tempo de associação com a população humana e o modelo preciso de transmissão. É importante ressaltar que um mesmo vírus pode apresentar diferentes padrões filogeográficos dependendo da escala espacial considerada (Holmes, 2008).

Atualmente, a filogeografia e a teoria da coalescência vêm sendo empregadas com sucesso para inferir a história da população de um patógeno e investigar a dinâmica evolutiva dos diferentes subtipos do HIV-1 nas diferentes regiões geográficas (Robbins, Lemey *et al.*, 2003; Salemi, De Oliveira *et al.*, 2005; Salemi, De Oliveira *et al.*, 2008; Salemi, Goodenow *et al.*, 2008; Gray, Tatem *et al.*, 2009).



**Figura 9. Representação esquemática de diferentes padrões filogeográficos exibidos por retrovírus que infectam humanos.** Cada população é representada por quadrados de diferentes cores e o tamanho dos quadrados representa o tamanho populacional. As árvores foram enraizadas de modo que o tempo evolutivo corre da esquerda (mais antigos) para a direita (mais recente). A. Ausência de estrutura espacial, com mistura randômica de populações. B. Transmissão do tipo onda, na qual o vírus se desloca de um ponto mais antigo para populações recente. C. Modelo fonte-dreno, no qual as populações em vermelho aparecem como fonte de vírus observado nas demais populações. D. Dinâmica gravitacional, em que os vírus movem das maiores populações antes de dispersarem para populações menores. E. Forte subdivisão espacial, na qual não há evidências de migrações entre populações.

## 2. Objetivos

### 2.1. Objetivo Geral

- Reconstruir a história e evolução dos subtipos B e C do HIV-1 no Brasil por meio de análises filogenéticas, de coalescência e filogeográficas.

### 2.2. Objetivos Específicos

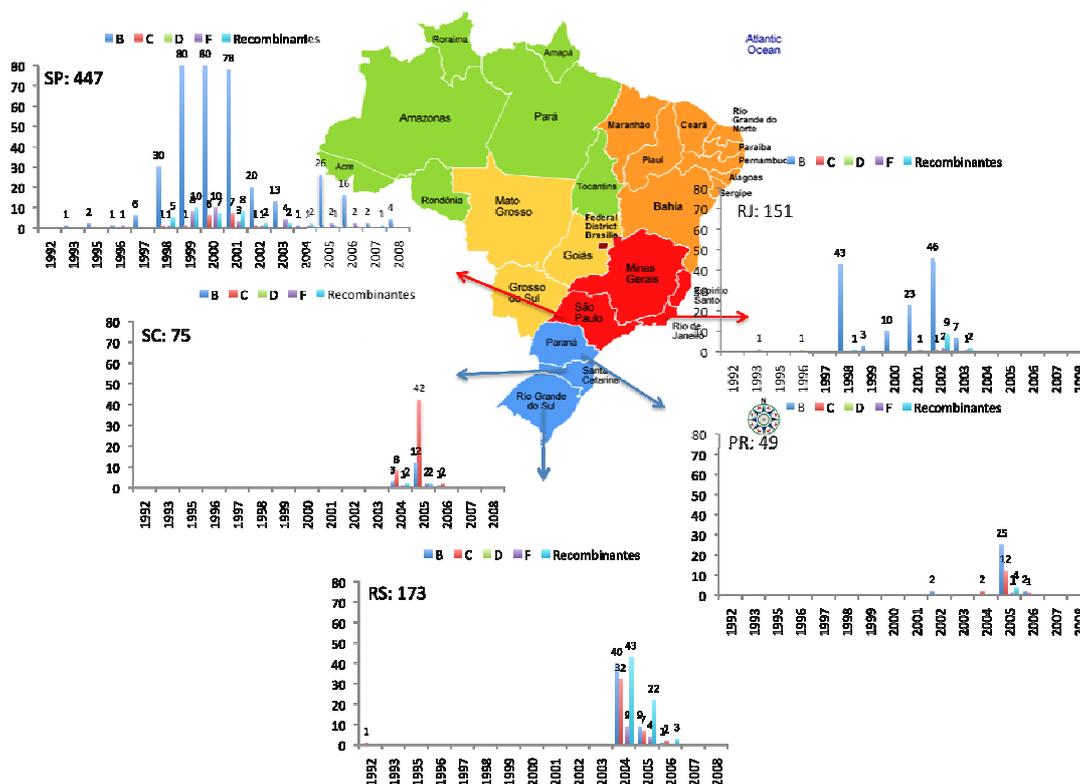
- Verificar a presença de mutações associadas à resistência em amostras de indivíduos não expostos a tratamento anti-retroviral no Brasil e relatar o efeito dessas mutações na suscetibilidade genética às diversas drogas anti-retrovirais;
- Verificar a variabilidade de subtipos do HIV-1 em indivíduos soropositivos no Brasil não expostos a tratamento anti-retroviral, por meio da análise de seqüências de *pol* depositadas nos bancos genômicos *Los Alamos HIV Sequence Database* (Los Alamos, 2009b) e *GenBank* (National Center for Biotechnology Information, 2009);
- Descrever o perfil genético de *pol* das formas recombinantes encontradas em amostras de indivíduos soropositivos no Brasil que não receberam tratamento anti-retroviral, de seqüências depositadas nos bancos genômicos *Los Alamos HIV Sequence Database* (Los Alamos, 2009b) e *GenBank* (National Center for Biotechnology Information, 2009), e compará-lo ao perfil das formas recombinantes circulantes (CRF) descritas na literatura;
- Verificar a presença de assinaturas na protease e transcriptase reversa que estejam associadas ou não a resistência aos anti-retrovirais em linhagens de HIV-1 isoladas de indivíduos no Brasil que não receberam tratamento anti-retroviral;
- Analisar regiões da protease e transcriptase reversa que codificam epitopos de importância no reconhecimento imunológico do HIV-1, em indivíduos soropositivos no Brasil que não usaram anti-retrovirais;
- Inferir a história demográfica dos subtipos B e C do HIV-1 no Brasil, por meio de análises de coalescência;
- Determinar as taxas evolutivas dos subtipos B e C do HIV-1 no Brasil;
- Estimar a data e local de entrada e as principais rotas de migração dos subtipos B e C do HIV-1 no Brasil.

### 3. Metodologia

#### 3.1. Amostragem

##### 3.1.1. Sequências genômicas

Primeiramente, com a finalidade de estudar a epidemia do HIV-1 no Brasil, foram selecionadas todas as seqüências brasileiras de *pol* (protease + transcriptase reversa) depositadas nos banco genômico *Los Alamos HIV Sequence Database* (Los Alamos, 2009b). No total, foram analisadas as seqüências dos genes da protease (nucleotídeos 2316 a 2504, em HXB2) e transcriptase reversa (nucleotídeos 2688 a 3158, em HXB2) de 895 amostras coletadas no Sudeste e Sul do Brasil no período de 1992 a 2008 (Figura 10). O Apêndice A relaciona cada isolado brasileiro analisado com o número de acesso do *GenBank*, ano da coleta, estado no qual a amostra foi coletada, e o artigo científico no qual foi publicado.

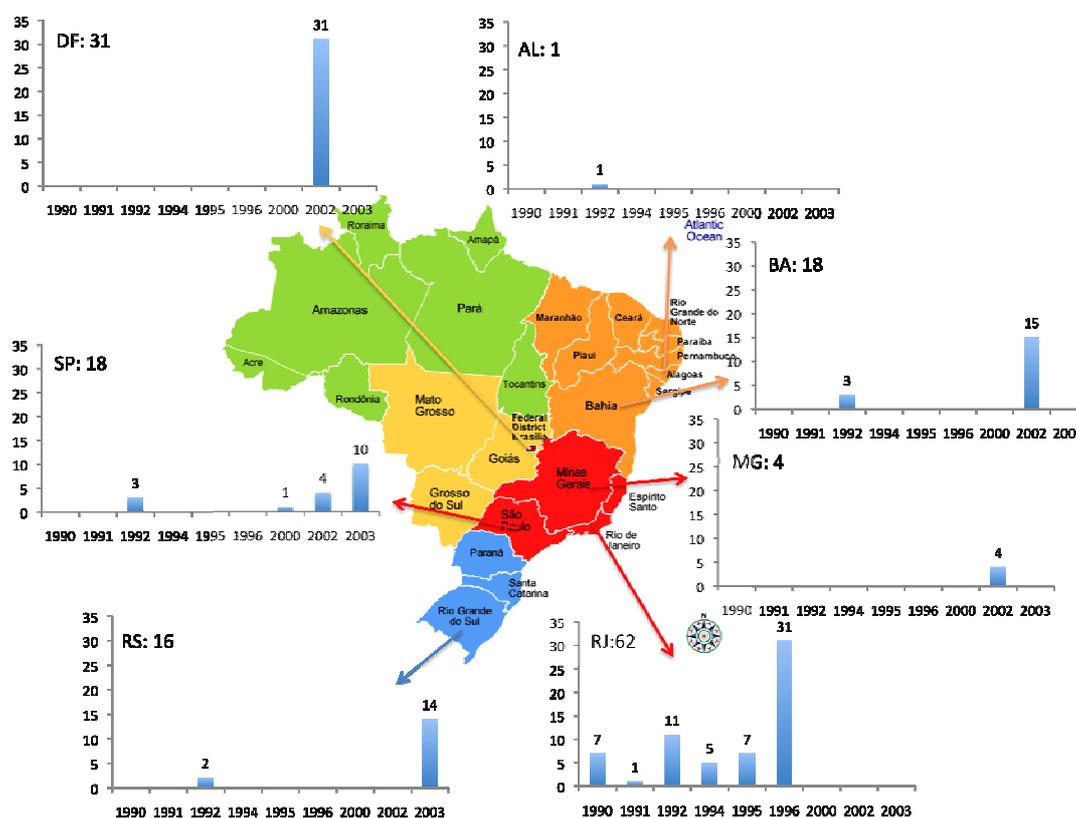


**Figura 10. Distribuição geográfica e cronológica das 895 seqüências de *pol* analisadas neste estudo.** Na porção esquerda superior de cada gráfico está identificado o estado representado pelo gráfico. PR, Paraná; RJ, Rio de Janeiro; RS, Rio Grande do Sul; SC, Santa Catarina e SP, São Paulo.

Para melhor caracterizar a relação entre as linhagens do subtipo B do HIV-1 circulantes no Brasil e aquelas prevalentes em outras regiões do mundo, foram selecionadas todas as seqüências brasileiras de *gp120* (*env*) previamente classificadas

como subtipo B, disponíveis no *Los Alamos HIV Sequence Database* (Los Alamos, 2009b). Além disso, foram selecionadas todas as seqüências de *pol* e *gp120* do subtipo B do HIV-1, provenientes da América do Norte, América Central, América do Sul, Europa, Ásia, África e Oceania, objetivando representar as linhagens do subtipo B circulantes nos diferentes continentes.

No total, foram analisadas 150 seqüências de *gp120* (nucleotídeos 7050 a 7295, em HXB2) do subtipo B do HIV-1 coletadas no Nordeste, Centro-Oeste, Sudeste e Sul do Brasil no período de 1990 a 2003 (Figura 11). O Apêndice B relaciona cada seqüência brasileira de *gp120* com o número de acesso do *GenBank*, ano da coleta, estado no qual a amostra foi coletada, motivo no *loop V3* de *gp120* e o artigo científico no qual foi publicado.

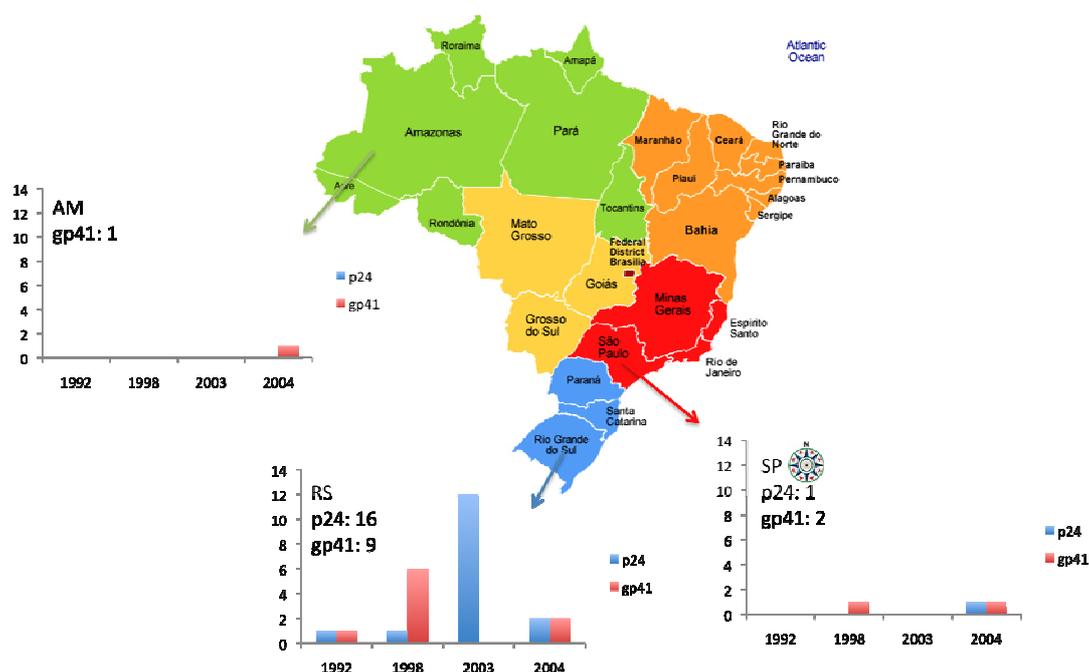


**Figura 11. Distribuição geográfica e cronológica das 150 seqüências de *gp120* (*env*) do subtipo B do HIV-1 analisadas neste estudo.** Na porção esquerda superior de cada gráfico está identificado o estado representado pelo gráfico. AL, Alagoas; BA, Bahia; DF, Distrito Federal; MG, Minas Gerais; RJ, Rio de Janeiro, RS, Rio Grande do Sul e SP, São Paulo.

Concernente ao estudo da epidemia do subtipo C do HIV-1 no Brasil, todas as seqüências brasileiras de *p24* (*gag*) e *gp41* (*env*) do subtipo C, disponíveis no *Los Alamos HIV Sequence Database* (Los Alamos, 2009b), foram selecionadas, assim

como as seqüências de *p24*, *pol* e *gp41* procedentes dos diferentes continentes, afim de se obter uma representação global desse subtipo.

Foram analisadas, ao todo, 17 seqüências de *p24* (nucleotídeos 1615 a 2041, em HXB2) e 13 de *gp41* (nucleotídeos 7769 a 8234, em HXB2) do subtipo C do HIV-1 coletadas no Norte, Sudeste e Sul do Brasil no período de 1991 a 2004 (Figura 12). O Apêndice C relaciona cada seqüência brasileira de *p24* e *gp41* com o número de acesso do *GenBank*, ano da coleta, estado no qual a amostra foi coletada e o artigo científico no qual foi publicado.



**Figura 12. Distribuição geográfica e cronológica das 17 seqüências de *p24* e 13 de *gp41* (*env*) do subtipo C do HIV-1 analisadas neste estudo.** Na porção esquerda superior de cada gráfico está identificado o estado representado pelo gráfico. AM, Amazonas; RS, Rio Grande do Sul e SP, São Paulo.

As seqüências incluídas no estudo obedeceram os seguintes critérios de inclusão: 1) ter sido publicada em periódicos científicos (com exceção das seqüências novas descritas no próximo item); 2) não serem de indivíduos relacionados epidemiologicamente; 3) apenas uma seqüência por paciente ser selecionada randomicamente; 4) o lugar e ano de coleta da amostra ser conhecido e claramente definido no artigo científico ou confirmado por um dos autores; 5) as seqüências de *pol* serem proveniente de indivíduos que não receberam tratamento anti-retroviral.

Este projeto foi aprovado pelo Comitê Científico e de Ética do Instituto Adolfo Lutz (IAL - BM - 09/08).

### 3.1.2. Técnicas de Processamento de Amostras para Seqüenciamento

Todos os experimentos concernentes a produção de novas seqüências de *pol*, e que, portanto, não foram publicadas previamente, foram realizados pela equipe do Laboratório de Genotipagem do Instituto Adolfo Lutz, em São Paulo. No total, foram seqüenciadas 118 amostras de São Paulo, 49 do Paraná e cinco de Santa Catarina.

#### 3.1.2.1 Extração do RNA Viral

As amostras de plasma foram descongeladas à temperatura ambiente e, em seguida, colocadas no *vortex* por cinco segundos, para misturar, e centrifugadas a 2.000 x *g* por dois segundos. Uma alíquota de 1mL de plasma foi centrifugada a 25.000 x *g* por duas horas a 4°C. Após a centrifugação, 900µL do sobrenadante foi removido e adicionou-se 300µl de Trizol® (Gibco/BRL, Life Technologies, USA) aos 100µL restantes de plasma. As amostras foram mantidas a temperatura ambiente por cinco minutos para garantir a lise completa do vírus. Foram, então, adicionados 80µL de clorofórmio a cada amostra e os tubos foram incubados por 15 minutos a temperatura ambiente. Procedeu-se a centrifugação a 12.000 x *g* por 15 minutos a 4°C, para que ocorresse a separação da fase aquosa que contém o RNA viral, e o sobrenadante foi transferido para um novo tubo. Em seguida, adicionou-se 200µL de isopropanol e 40µL de dextran T500 a 1µg/µL, seguindo de nova centrifugação a 12.000 x *g* por 10 minutos a 4°C, para que ocorresse a precipitação do material genético. O sobrenadante foi desprezado e o precipitado lavado com 400µL de etanol 75%, por meio de centrifugação a 7.500 x *g* por 5 minutos a 4°C. Novamente, removeu-se o máximo possível de sobrenadante e o precipitado foi ressuspenso em 50µl de água *DNase RNase free* (Gibco/BRL, Life Technologies, USA). O RNA viral foi armazenado a -80° C.

#### 3.1.2.2. Transcrição Reversa

A reação de transcrição reversa foi realizada em um volume final de 20µL. Inicialmente, uma solução contendo 10,5µL de RNA viral, 300ng de iniciadores randômicos (Invitrogen, USA), 500µM de dNTP (Invitrogen, USA) e 1,0µL de água Milli Q foi mantida a 65°C por cinco minutos. Em seguida, foi resfriada em gelo por cinco minutos para manter a estrutura relaxada. Adicionou-se, então, 50mM de Tris-HCl pH8,3 (Invitrogen, USA), 75mM de KCl (Invitrogen, USA), 3mM de MgCl<sub>2</sub> (Invitrogen, USA), 250mM de DTT (Invitrogen, USA), 10 unidades (U) de inibidor de RNase (Invitrogen, USA) e 100U da enzima transcriptase reversa Super Script III

(Invitrogen, USA). Essa reação foi incubada a 25°C por cinco minutos e, em seguida, a 50°C por 60 minutos, sendo inativada por aquecimento a 70°C por 15 minutos. O cDNA sintetizado foi armazenado a -20°C.

### 3.1.2.3. Reação em Cadeia da Polimerase (PCR)

O DNA complementar (cDNA) do HIV-1 foi amplificado por meio de *nested PCR*. O primeiro ciclo foi realizado num volume final de 50µL, contendo 5µL de cDNA, 20mM de Tris-HCl pH8,4 (Invitrogen, USA), 50mM de KCl (Invitrogen, USA), 2,5mM de MgCl<sub>2</sub> (Invitrogen, USA), 400uM de dNTP (Invitrogen, USA), 300nM dos iniciadores K1 e K2 (Rodrigues, Custodio *et al.*, 2005), 1,25U de *Taq* Polimerase Platinum® (Invitrogen, USA) e água *DNase RNase free* (Gibco/BRL, Life Technologies, USA). A amplificação em termociclador foi programada para 94°C por três minutos, seguida de 35 ciclos de (94°C por 30 segundos, 55°C por 30 segundos e 72°C por dois minutos), e extensão final de 72°C por 10 minutos. Essa reação permitiu a obtenção de um fragmento de DNA de 1,2kb do gene *pol*.

O segundo ciclo foi realizado num volume final de 50µL, contendo 2µL da primeira reação, 20mM de Tris-HCl pH8,4 (Invitrogen, USA), 50mM de KCl (Invitrogen, USA), 2,5mM de MgCl<sub>2</sub> (Invitrogen, USA), 400uM de dNTP (Invitrogen, USA), 200nM de iniciadores, 1,25U de *Taq* Polimerase Platinum® (Invitrogen, USA) e água *DNase RNase free* (Gibco/BRL, Life Technologies, USA).

Para a amplificação do gene da protease, foram utilizados os iniciadores DP10 e DP11 (Rodrigues, Custodio *et al.*, 2005) e a amplificação em termociclador foi programada para 94°C por três minutos, seguida de 25 ciclos de (94°C por 30 segundos, 55°C por 30 segundos e 72°C por um minuto), e extensão final de 72°C por 10 minutos.

Para a obtenção de parte da região da transcriptase reversa, foram utilizados os iniciadores F1 e F2 (Rodrigues, Custodio *et al.*, 2005) e o seguinte programa de ciclagem: 94°C por 3 minutos, seguida de 35 ciclos de (94°C por 1 minuto, 55°C por 30 segundos e 72°C por 1 minuto e 30 segundos), e extensão final de 72°C por 10 minutos.

Controles negativos foram incluídos durante os procedimentos de amplificação para monitorar possíveis contaminações. Todos os iniciadores estão descritos na Tabela 1.

**Tabela 1. Lista de iniciadores utilizados para amplificar e seqüenciar os genes da protease (PR) e transcriptase reversa (RT).**

Iniciador	Seqüência (5' - 3')	Gene	Localização <sup>a</sup>
K1	CAGAGCCAACAGCCCCACCA	<i>pol</i>	2147-2166
K2	TTTCCCCACTAACTTCTGTATGTCATTGACA	<i>pol</i>	3338-3308
DP10	TAACTCCCTCTCAGAAGCAGGAGCCG	PR	2198-2223
DP 11	CCATTCTGGCTTTAATTTTACTGGTA	PR	2598-2572
F1	GTTGACTCAGATTGGTTGCAC	RT	2519-2539
F2	GTATGTCATTGACAGTCCAGC	RT	3321-3301

<sup>a</sup> relativa à linhagem de referência do HIV-1 HXB2.

#### 3.1.2.4. Detecção, Quantificação e Purificação do Produto de PCR

As amostras amplificadas foram submetidas à eletroforese em gel de agarose a 1% em tampão TBE 0,5X, corado com 0,5µg de brometo de etídeo.

Os produtos da amplificação e o padrão de tamanho molecular de 100pb (Invitrogen,USA), diluído 1:10, foram aplicados no gel de agarose e submetidos a uma corrente constante de 100v, por 30 minutos.

As amostras identificadas como positivas foram submetidas a um processo de purificação utilizando colunas de sílica (Rapid Concert, Gibco/BRL, Life Technologies, USA). O material purificado foi quantificado em gel de agarose a 1% em tampão TBE 0,5X e corado com 0,5µg de brometo de etídeo, com padrão de massa molecular *Low Mass Ladder* (Invitrogen, USA).

#### 3.1.2.5. Reação de Seqüenciamento

Para a reação de seqüenciamento foi utilizado *DNA Sequencing Kit –Big Dye™ Terminator Cycle Sequencing Ready Reaction – ABI Prism®* (Applied Biosystems, USA) e de 5 a 10ng de DNA amplificado.

Para a região da protease, a reação foi realizada utilizando-se 2µL de *Big Dye* (Applied Biosystems), 2,5x tampão Tris-HCl e 320nM dos iniciadores DP10 e DP11, para um volume final de 20µL.

Para a região da transcriptase reversa, utilizou-se 4µL de *Big Dye* (Applied Biosystems), 2,5x tampão Tris-HCl e 320nM de iniciadores F1 e F2, para um volume final de reação de 20µL.

Todas as reações de incorporação de rodaminas foram realizadas em placas de 96 poços (*Optical 96-well Reaction Plate*, Applied Biosystems), aplicando-se 25 ciclos de (96°C por 10 segundos, 50°C por cinco segundos e 6°C por quatro minutos).

### **3.1.2.6. Purificação da Reação de Seqüenciamento**

Foram adicionados em cada poço da placa 80µL de isopropanol. Em seguida, a placa foi centrifugada a 1.990 x g por 45 minutos; o isopropanol foi desprezado por inversão da placa e prosseguiu-se uma etapa de centrifugação da placa invertida a 150 x g por um minuto. O precipitado seco foi ressuspendido em 10µL de formamida Hi-Di e as amostras foram seqüenciadas em Seqüenciador Automático ABI 3100.

### **3.1.2.7. Edição das Seqüências**

Os cromatogramas foram alinhados e editados manualmente, utilizando os programas Sequence Navigator (Applied Biosystems, USA) e Sequencher (Gene Codes, USA). A edição automatizada, realizada a partir dos cromatogramas, foi feita utilizando o programa Phred ([www.ial.sp.gov.br](http://www.ial.sp.gov.br)).

## **3.2. Detecção de Mutações Associadas à Resistência e Análise da Susceptibilidade aos Anti-retrovirais**

A presença de mutações associadas à resistência a diferentes drogas anti-retrovirais foi definida por meio de um algoritmo computadorizado, o programa HIVdb (Stanford University, 2009). Esse programa atribui uma pontuação para cada mutação em relação a cada droga. A pontuação total para um anti-retroviral é obtida pela adição das pontuações associadas a cada mutação. Com isso, o programa utiliza a pontuação total de cada droga para inferir um dos seguintes níveis de resistência: susceptível, baixo nível de resistência potencial, baixo nível de resistência, resistência intermediária e alto nível de resistência.

Neste estudo, os níveis de resistência foram classificados da seguinte maneira: amostras susceptíveis (pontuação 0 a 14), amostras com baixo nível de resistência ou com resistência intermediária foram consideradas como possivelmente resistentes (resistência intermediária, pontuação 15 a 29) e amostras com alto nível de resistência foram consideradas resistentes (pontuação acima de 30). Assim, pôde-se obter uma análise da susceptibilidade ou do nível de resistência da seqüência submetida em relação às drogas anti-retrovirais.

### 3.3. Mapeamento de Verossimilhança

O sinal filogenético das seqüências de nucleotídeos alinhadas foi verificado por mapeamento de verossimilhança. Nessa análise, verifica-se a probabilidade das três possíveis topologias que podem ocorrer para o conjunto de quatro seqüências, escolhidas aleatoriamente, denominado de quarteto. Para isso emprega-se o método de máxima verossimilhança e cada probabilidade calculada é marcada como um ponto em um triângulo equilátero (mapa de verossimilhança). Os pontos localizados próximos aos vértices do triângulo representam os sinais filogenéticos típicos de árvores, indicando topologia totalmente resolvida; os situados no centro representam filogenias em forma de estrela e os nas laterais indicam filogenia em forma de rede, resultante da presença de recombinação, por exemplo. A distribuição dos pontos em cada área do triângulo fornece uma idéia de como as linhagens analisadas evoluíram (Strimmer e Von Haeseler, 1997). O ideal para a realização de análise filogenética é que a soma dos três vértices do triângulo seja o mais próximo possível de 100% (Schmidt e Von Haeseler, 2009). O mapeamento de verossimilhança foi realizado utilizando-se o programa TREE-PUZZLE (Schmidt, Strimmer *et al.*, 2002).

### 3.4. Análises Filogenéticas

Para determinar o subtipo das seqüências de *pol*, assim como confirmar o das seqüências de subtipo B e C, conforme determinado no *Los Alamos HIV Sequence database* (Los Alamos, 2009b), as seqüências estudadas por esse trabalho foram alinhadas com as seguintes referências: subtipo A1 (AF004885, AF069670), subtipo A2 (AF286237, AF286238) subtipo B (U21135, AY173951, AY423387, AY331295), subtipo C (AF67155, AY772699, U46016), subtipo D (K03454, U88824, AY253311, AY371157), subtipo F1 (AF077336, AF075703, AF005494, AJ249238), subtipo F2 (AY371158, AF377956, AJ249237, AJ249236), subtipo G (AF061642, U88826, AF061641, AF084936), subtipo H (AF190128, AF005496, AF190127), subtipo J (AF082395, AF082394) e subtipo K (AJ249239, AJ249235). Para verificar a relação das seqüências de subtipo B e C circulantes no Brasil com aquelas desses subtipos prevalentes em outros continentes, foram alinhadas as seqüências estudadas por esse trabalho; seqüências de referência dos subtipos B e C, respectivamente, de diferentes países, selecionadas aleatoriamente; e as seqüências A1 (AF004885, AF069670), C (AF67155, AY772699) e J (AF082395, AF082394), para análises do subtipo B, e A1 (AF004885, AF069670), B (U21135, AY423387) e J (AF082395, AF082394), para análises do subtipo C, utilizadas como grupo externo.

Os alinhamentos foram conduzidos no programa Clustal W (Thompson, Higgins *et al.*, 1994) e editadas no BioEdit (Hall, 2005). Durante a edição, foram retiradas regiões em que mais da metade das amostras apresentavam deleções e os excessos das extremidades 5' e 3'. Desse modo, todas as seqüências ficaram com o mesmo tamanho e manteve-se a fase de leitura.

Árvores filogenéticas de máxima verossimilhança (MV) foram inferidas pelo programa PhyML (Guindon e Gascuel, 2003), disponível em <http://www.atgc-montpellier.fr/phyml/>, utilizando o modelo de substituição nucleotídica GTR, incluindo o parâmetro de distribuição no formato Gamma e sítios invariáveis, de acordo com o teste de verossimilhança hierárquica descrito por Swofford e Sullivan (2009) (Swofford e Sullivan, 2009). A confiabilidade das árvores foi avaliada pelo teste de razão de verossimilhança aproximada tipo SH (aLRT) (Anisimova e Gascuel, 2006), o qual compara a verossimilhança da melhor e da segunda melhor alternativa de agrupamento dos ramos de interesse na árvore filogenética. De acordo com a análise da taxa de erro do tipo I (teste significativo | ramo não correto), a aLRT de um ramo interno é praticamente exato para um valor igual ou superior a 0,9. As árvores foram visualizadas e editadas no programa FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 3.5. Análise de Recombinação

Além das análises filogenéticas, a presença de seqüências recombinantes do HIV-1 foram também analisadas pelo programa *Recombination Detection Program* – RDP, versão 3.41 (Martin, Williamson *et al.*, 2005). Para detecção de recombinação foram utilizados os métodos RDP, GENECONV, Chimaera, Maxchi e 3Seq. Para confirmar os sinais de recombinação identificados, foram utilizados os métodos Bootscan e Siscan. Só foram considerados os sinais de recombinação detectados por mais de um método. Os parâmetros utilizados em cada método foram:

- RDP
  - ✓ selecionar apenas seqüências de referência internas.
  - ✓ janela: 30pb.
  - ✓ detectar recombinação entre seqüências com 0% a 100% de identidade.
- GENECONV
  - ✓ Tratar blocos de deleções e inserções como um polimorfismo.
  - ✓ Escala G: 1.

- ✓ Tamanho mínimo de fragmento alinhado: 1.
- ✓ Número mínimo de polimorfismos no fragmento: 2.
- ✓ Pontuação mínima de pareamento no fragmento: 2.
- ✓ Número máximo de fragmentos sobrepostos: 1.
- Chimaera
  - ✓ janela: tamanho variável.
  - ✓ fração de sítios variáveis por janela: 0,1.
- Maxchi
  - ✓ janela: tamanho variável.
  - ✓ fração de sítios variáveis por janela: 0,1.
  - ✓ Strip Gap = on.
- Bootscan
  - ✓ janela: 350pb
  - ✓ passo: 10pb
  - ✓ Usar árvores Neighbour-joining
  - ✓ Número de *bootstrap*: 100
  - ✓ Modelo: Felsenstein, 1984
- Siscan
  - ✓ janela: 350pb
  - ✓ passo: 10pb
  - ✓ Strip Gap = on
  - ✓ número de permutações:100

A localização dos pontos de recombinação foi confirmada pelas matrizes Maxchi e LARD. As regiões entre os pontos de recombinação foram analisadas separadamente por filogenia utilizando-se a mesma metodologia descrita anteriormente.

### **3.6. Análise da Presença de Assinaturas na Protease e Transcriptase Reversa**

Seqüências dos subtipos B, C e F das formas recombinantes foram alinhadas com as seqüências consenso dos subtipos B, C e F pelo programa Clustal W (Thompson, Higgins *et al.*, 1994), prosseguindo-se com a edição de acordo com o descrito previamente. A seqüência de aminoácidos foi deduzida utilizando-se o programa BioEdit e a presença de polimorfismo foi avaliada de acordo com a seqüência consenso do subtipo B. Foram destacados os polimorfismos que apareceram em, pelo menos, 50% dos isolados. As mutações encontradas em, pelo

menos, 80% das seqüências de um determinado subtipo, e não encontrada na seqüência consenso do subtipo analisado, foi considerada como provável assinatura do subtipo em questão, circulante nas regiões geográficas analisadas. Em seguida, foi verificado se as assinaturas descritas neste estudo apareceram nas formas recombinantes analisadas neste trabalho.

### 3.7. Identificação de Epitopos na Protease e Transcriptase Reversa

As 895 amostras analisadas por esse estudo tiveram suas seqüências de aminoácidos deduzidas a partir das seqüências de nucleotídeos. Em seguida, foram alinhadas com a seqüência de referência do subtipo B, HXB2, utilizando-se o programa CLUSTAL W (Thompson, Higgins *et al.*, 1994).

Os epitopos reconhecidos por anticorpos e células CD4<sup>+</sup> foram identificados de acordo com a sua localização em HXB2, baseando-se na lista de epitopos disponível no *Los Alamos HIV Immunology Database* (Los Alamos, 2009a). A identificação de epitopos reconhecidos por células CD8<sup>+</sup> foi baseada na lista de epitopos de linfócitos T citotóxicos derivados do HIV ótimos para o estudo de evolução viral descrita por Frahm *et al* (2008) (Frahm, Baker *et al.*, 2008).

### 3.8. Análises de Relógio Molecular

A história demográfica da epidemia dos subtipos B e C do HIV-1 no Brasil foi inferida a partir da relação genealógica das linhagens brasileiras que formaram um clado monofilético, baseando-se na teoria da coalescência (Kingman, 2000). Foram considerados seis modelos demográficos para a epidemia de cada genótipo: tamanho populacional constante com relógio molecular rígido ou relaxado, crescimento exponencial com relógio molecular rígido ou relaxado e plotagem Bayesiana (*Bayesian skyline plot* – BSP) com relógio molecular rígido ou relaxado.

A taxa evolutiva (número de substituições nucleotídicas por sítio por ano) e o tempo do ancestral comum mais recente (Tmrca, anos) foram estimados aplicando-se relógio molecular relaxado e os modelos demográficos de tamanho populacional constante, crescimento exponencial e BSP. Foram conduzidas análises separadas utilizando como premissas taxa evolutiva uniforme ou normal, com média e desvio padrão escolhido de acordo com estimativas previamente publicadas:  $13 \pm 5 \times 10^{-4}$  para p24 (Lemey, Pybus *et al.*, 2004),  $16 \pm 6 \times 10^{-4}$  para RT (Bello, Passaes *et al.*, 2008),  $5 \pm 2 \times 10^{-3}$  para gp120 (Bello, Eyer-Silva *et al.*, 2007) e  $31 \pm 9 \times 10^{-4}$  para gp41 (Lemey, Pybus *et al.*, 2004). Nessas análises, além das seqüências brasileiras, foram incluídas seqüências de referência de cada subtipo do HIV-1: subtipo A1 (AF004885,

AF069670), subtipo A2 (AF286237, AF286238) subtipo B (U21135, AY173951, AY423387, AY331295), subtipo C (AF67155, AY772699, U46016), subtipo D (K03454, U88824, AY253311, AY371157), subtipo F1 (AF077336, AF075703, AF005494, AJ249238), subtipo F2 (AY371158, AF377956, AJ249237, AJ249236), subtipo G (AF061642, U88826, AF061641, AF084936), subtipo H (AF190128, AF005496, AF190127), subtipo J (AF082395, AF082394) e subtipo K (AJ249239, AJ249235).

Todas as estimativas Bayesianas foram realizadas no programa BEAST versão 1.4.8 (Drummond e Rambaut, 2007) empregando o método de relógio Bayesiano Monte Carlo de Cadeias de Markov (*Bayesian Markov Chains Monte Carlo* – MCMC) para 100.000.000 gerações com amostragem a cada 10.000 gerações, utilizando-se o modelo evolutivo GTR, para o subtipo B e HKY, para o C. A representatividade das cadeias de Markov foi verificada pelo valor de tamanho amostral efetivo (TAE) maior que 200. A comparação dos diferentes modelos demográficos foi realizada por meio do Fator Bayes (FB), calculado pelo programa BEAST versão 1.4.8 (Drummond e Rambaut, 2007), de acordo com as instruções do programa encontradas em: [http://beast.bio.ed.ac.uk/Model\\_comparison](http://beast.bio.ed.ac.uk/Model_comparison). O FB é a razão das verossimilhanças marginais (marginal em relação a premissa) de dois modelos a serem comparados. Os cálculos de verossimilhança marginal aproximada foi realizada para cada modelo coalescente por meio de 1000 réplicas (*bootstrap*) utilizando a média harmônica das verossimilhanças amostradas. A diferença (em  $\log_e$ ) da verossimilhança marginal entre dois modelos é igual ao  $\log_e$  do Fator de Bayes,  $\log_e(\text{FB})$ . Evidências contra a hipótese nula é indicado por  $2 \cdot \log_e(\text{FB}) > 3$  (moderado) e  $> 10$  (forte). Os cálculos do FB foram realizados no programa Tracer v. 1.4.1 (<http://beast.bio.ed.ac.uk/Tracer>). As árvores de máxima probabilidade posterior (MAP) foram selecionadas da distribuição posterior de árvores pelo programa TreeAnnotator v. 1.4.8 disponível no pacote de programas BEAST. As árvores foram visualizadas no programa FigTree v.1.2.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 3.9. Análises Filogeográficas

A migração de/para de diferentes subpopulações virais dos subtipo B e C do HIV-1 o Brasil foi testada por uma versão modificada do teste Slatkin e Maddison (Slatkin, 1989; Salemi, Lamers *et al.*, 2005), usando o programa MacClade versão 4.06 (Maddison e Maddison, 2008). Para essa análise, foram incluídas apenas as seqüências brasileiras que formaram um clado monofilético.

Para a análise do subtipo B, foram analisados cinco alinhamentos de gp120 contendo, cada um, duas seqüências selecionadas aleatoriamente do Distrito Federal, Rio de Janeiro, São Paulo e Rio Grande do Sul, além de uma seqüência da Bahia e uma de Santa Catarina. Por sua vez, a migração do subtipo C no Brasil foi avaliada por quatro alinhamentos cada um contendo 15 seqüências selecionadas randomicamente de cada um dos seguintes estados: São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. O número de seqüências representando cada estado foi semelhante para evitar viés amostral nas estimativas de fluxo gênico devido a diferenças grandes no tamanho amostral de diferentes sub-populações (Slatkin, 1989; Salemi, Lamers *et al.*, 2005).

Uma matriz de um caractere foi obtida do alinhamento, assumindo para cada taxa na árvore um código de uma letra indicando a região geográfica de origem. A origem putativa de cada seqüência ancestral na árvore foi então inferida por meio da reconstrução mais parcimoniosa (RMP) do ancestral do caractere. A filogenia utilizada por esta análise foi procedente da árvore MAP obtida por inferência Bayesiana com o programa BEAST v.1.4.8 (Drummond e Rambaut, 2007). O número de migrações observadas na genealogia foi computada e comparada àquela obtida por uma distribuição de 10.000 árvores randômicas obtidas pelo método *joining-splitting*. A observação de genealogias com menor número de migrações significativo indica a presença de populações subdivididas com fluxo gênico restrito. Migrações específicas entre diferentes regiões geográficas foram traçadas pela ferramenta *State change and statis* (MacClade software), o qual contabiliza o número de mudanças em uma árvore para cada par de caracteres. A matriz de migração resultante foi normalizada e foi realizado o teste de randomização com 10.000 matrizes obtidas de 10.000 árvores randômicas para verificar a significância estatística do número de migrações observadas. A tendência das rotas de migração foi definido pelo teste de metapopulação. Nesse teste, foi calculada a porcentagem de migrações do estado A para o B menos a porcentagem de migrações do estado B para o A, definindo, portanto, as principais rotas de fluxo gênico para cada alinhamento avaliado. Em seguida, foi realizada a média da porcentagem de migração de cada rota de fluxo gênico, considerando os diferentes alinhamentos analisados para cada subtipo. Desse modo, pode-se determinar a tendência da direção e porcentagem de migração dos subtipo B e C do HIV-1 nos diferentes estados brasileiros analisados.

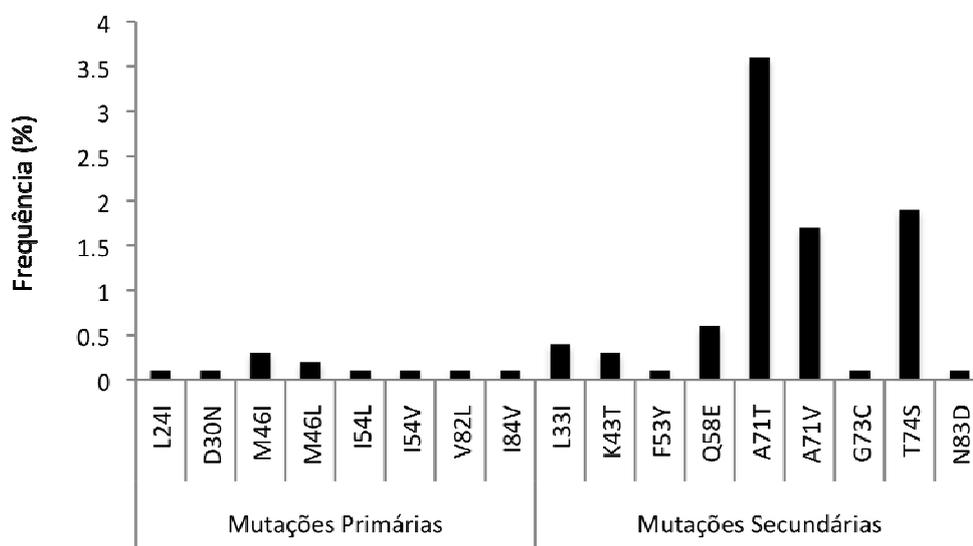
## 4. Resultados

### 4.1. Detecção de Mutações Associadas à Resistência e Análise da Susceptibilidade aos Anti-retrovirais

As 895 seqüências de *pol* do HIV-1 isolado de pacientes não expostos a tratamento anti-retroviral do Sul e Sudeste do Brasil foram analisadas pelo programa HIVdb (Liu e Shafer, 2006), disponível em *HIV drug resistance database* (Stanford University, 2009).

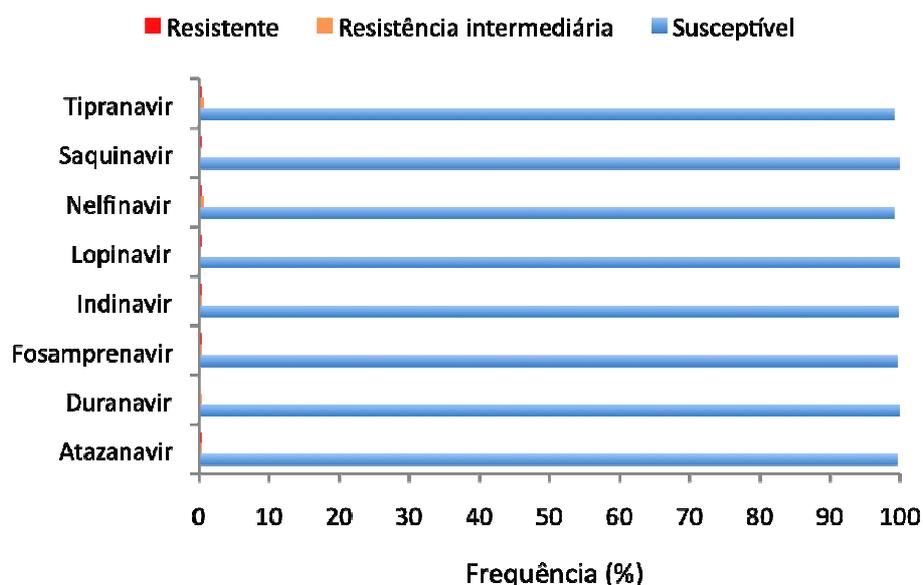
A análise das seqüências deduzidas de aminoácidos da protease (aminoácidos 22 a 84 da protease, em HXB2) revelou a presença de oito mutações primárias nas posições 24 (L24I), 30 (D30N), 46 (M46I e M46L), 54 (I54L e I54V), 82 (V82L) e 84 (I84V). As mutações L24I, D30N, I54L, I54V, V82L e I84V foram encontradas em uma (0,1%) amostra e as do códon 46, M46I e M46L, em três (0,3%) e duas (0,2%) amostras, respectivamente (Figura 13).

Foram, ainda, identificadas nove mutações secundárias na protease: F53Y, G73C e N83D em uma (0,1%) amostra; K43T em três (0,3%) amostras; L33I em quatro (0,4%) amostras; Q58E em cinco (0,6%); A71V em 15 (1,7%) amostras; T74S em 17 (1,9%) amostras e A71T em 32 (3,6%) amostras (Figura 13).



**Figura 13.** Frequência de mutações primárias e secundárias associadas à resistência aos inibidores de protease em isolados de HIV-1 de pacientes não expostos a tratamento anti-retroviral, do Sudeste e Sul do Brasil.

Quanto à resistência aos PI, uma (0,1%) amostra (DW4SP08) apresentou vírus resistente ao ATV, FPV, IDV, LPV e SQV. Três (0,3%) amostras (AW8SP01, DI6SP03 e DW4SP01) foram classificadas como resistentes ao NFV e duas (0,2%), DW4SP01 e DX8SP04, ao TPV . Além disso, duas (0,2%) amostras apresentaram resistência intermediária ao ATV, AW8SP01 e CL7RJ01, e ao FPV, AW8SP01 e BT7RS04. Uma (0,1%) amostra apresentou resistência intermediária ao DRV, BT7RS04, e ao IDV, AW8SP01. Cinco (0,6%) amostras (AH6SP99, AH9SP99, AN7SP00, BI4RS04 e BT7RS04) possuíam resistência intermediária ao NFV e seis (0,7%), BL3RS04, CU4RJ02, CZ7RS04, CZ8RS04, DG2RS04 e DU4SP02, ao TPV (Figura 14). De um modo geral, quatro (0,4%) amostras foram classificadas como resistentes e 13 (1,4%) apresentaram resistência intermediária a pelo menos um PI.



**Figura 14.** Níveis de resistência de isolados do HIV-1 de pacientes não expostos a anti-retrovirais aos inibidores de protease no Sudeste e Sul do Brasil.

O Quadro 1 relaciona as seqüências que apresentaram mutações primárias e/ou secundárias na protease com o ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos PI.

**Quadro 1. Descrição das seqüências que apresentaram mutações primárias e/ou secundárias na protease relacionando-os ao ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos inibidores de protease.**

Seqüência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de protease																
				Mutações primárias								Mutações secundárias								
				L 24 I	D 30 N	M 46 I	M 46 L	I 54 L	I 54 V	V 82 L	I 84 V	L 33 I	K 43 T	F 53 Y	Q 58 E	A 71 T	A 71 V	G 73 C	T 74 S	N 83 D
AA2SP99	1999	SP	B												X					
AA5SP99	1999	SP	B										X							
AA9SP99	1999	SP	Recombinante															X		
AC6SP98	1998	SP	Recombinante											X						
AD8SP99	1999	SP	B											X						
AE4SP99	1999	SP	B											X						
AF4SP99	1999	SP	B											X						
AJ1SP99	1999	SP	C															X		
AH6SP99	1999	SP	B			X														
AH9SP99	1999	SP	B				X													
AL5SP98	1998	SP	B											X						
AN7SP00	2000	SP	B				X													
AO3SP00	2000	SP	B											X						
AQ5SP00	2000	SP	B												X					
AQ8SP98	1998	SP	B											X						
AS2SP00	2000	SP	B								X									
AT8SP00	2000	SP	B											X						
AW8SP01	2001	SP	B	X		X														
BB8SP01	2001	SP	B															X		
BC9SP01	2001	SP	B											X						
BF3SP02	2002	SP	B											X						
BG5SP98	1998	SP	B										X							
BH7SP01	2001	SP	B											X						

Quadro 1. Continuação.

Sequência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de protease															
				Mutações primárias								Mutações secundárias							
				L 24 I	D 30 N	M 46 I	M 46 L	I 54 L	I 54 V	V 82 L	I 84 V	L 33 I	K 43 T	F 53 Y	Q 58 E	A 71 T	A 71 V	G 73 C	T 74 S
BI4RS04	2004	RS	B													X	X		
BJ3RS04	2004	RS	B											X					
BL3RS04	2004	RS	B									X							
BL7RS04	2004	RS	B										X						
BM2RS04	2004	RS	B												X				
BR7RS04	2004	RS	Recombinante															X	
BS1RS06	2006	RS	Recombinante															X	
BS3RS04	2004	RS	C															X	
BT7RS04	2004	RS	C					X											
BU3RS05	2005	RS	B												X				
BZ7SC05	2005	SC	C															X	
CA8SC05	2005	SC	C															X	
CB3SC05	2005	SC	C															X	
CD9RJ96	1996	RJ	D												X				
CE9RJ98	1998	RJ	B											X					
CF1RJ98	1998	RJ	B												X				
CF2RJ98	1998	RJ	B											X					
CG6RJ98	1998	RJ	B											X					
CH5RJ98	1998	RJ	B											X					
CH8RJ98	1998	RJ	B											X					
CI1RJ98	1998	RJ	B									X							
CI6RJ98	1998	RJ	B											X					
CL5RJ01	2001	RJ	B											X					

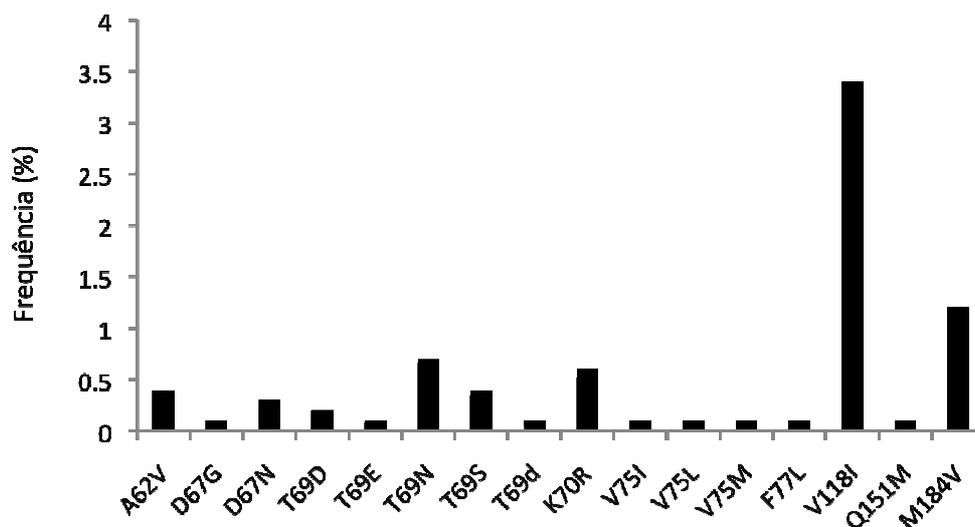


Quadro 1. Continuação.

Sequência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de protease															
				Mutações primárias								Mutações secundárias							
				L 24 I	D 30 N	M 46 I	M 46 L	I 54 L	I 54 V	V 82 L	I 84 V	L 33 I	K 43 T	F 53 Y	Q 58 E	A 71 T	A 71 V	G 73 C	T 74 S
DP1PR06	2006	PR	C																X
DU3SP02	2002	SP	B																
DU4SP02	2002	SP	B																
DV3SP03	2003	SP	B																
DW3SP00	2000	SP	B																
DW4SP08	2008	SP	B			X			X		X								
DW5SP08	2008	SP	B																
DX7SP04	2004	SP	Recombinante																
DX8SP04	2004	SP	B							X									
DY8SP05	2005	SP	B																
EA3SP05	2005	SP	B																
EB1SP05	2005	SP	B																
EB4SP06	2006	SP	B																
EC5SC05	2005	SC	C																

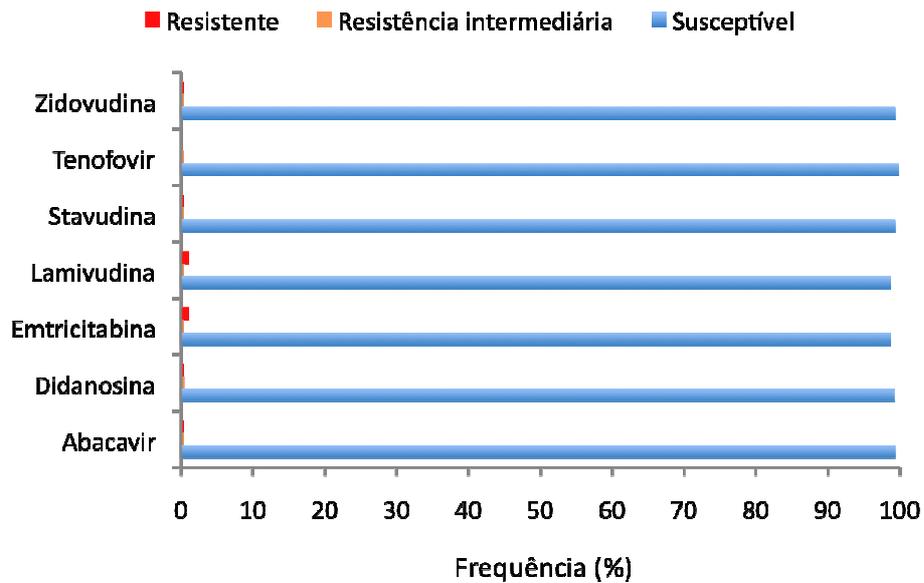
PR, Paraná; RJ, Rio de Janeiro; RS, Rio Grande do Sul; SC, Santa Catarina; e SP, São Paulo.

A análise das seqüências deduzidas de aminoácidos da transcriptase reversa (amoniácidos 47 a 203 da transcriptase reversa, em HXB2) revelou a presença de 16 mutações que diminuem a susceptibilidade aos NRTI: A62V e T69S em quatro (0,4%) amostras; D67G, T69E, T69d, V75I, V75L, V75M, F77L e Q151M em uma (0,1%); D67N em três (0,3%); T69N em seis (0,7%); K70R em cinco (0,6%); V118I em 30 (3,4%) e M184V em 11 (1,2%) (Figura 15).



**Figura 15.** Frequência de mutações associadas à resistência aos inibidores de transcriptase reversa análogos de nucleosídeos em isolados de HIV-1 de pacientes não expostos a anti-retrovirais, do Sudeste e Sul do Brasil.

Quanto à resistência aos NRTI, duas (0,2%) amostras (AZ6SP01 e BI2RS04) apresentaram vírus resistente ao ABC e DDI; 10 (1,1%) amostras (AX1SP01, BI2RS04, BI3RS04, BI7RS04, BZ5SC04, BZ7SC05, CC1SC05, DN4PR05, DW4SP08 e EC5SC05) a FTC e 3TC; e três (0,3%) amostras a D4T (AZ6SP01, BI2RS04 e DH6SP99) e AZT (AZ6SP01, BI4RS04 e DH6SP99). Além disso, três (0,3%) amostras, BZ5SC05, DH6SP99 e DW4SP08, apresentaram resistência intermediária a ABC; quatro (0,4%) a DDI (BZ5SC05, CU1RJ02, DH6SP99 e DW4SP08) e duas (0,2%) a FTC (AZ6SP01 e DH6SP99), 3TC (AZ6SP01 e DH6SP99), D4T (AR1SP00 e BI4RS04), FDT (AZ6SP01 e DH6SP99) e AZT (AR1SP00 e BI2RS04) (Figura 16). Dessa forma, 13 (1,4%) amostras foram classificadas como resistentes e oito (0,9%) apresentaram resistência intermediária a pelo menos um NRTI.



**Figura 16. Níveis de resistência de isolados do HIV-1, de pacientes não expostos a anti-retrovirais, aos inibidores de transcriptase reversa análogos de nucleosídeos no Sudeste e Sul do Brasil.**

O Quadro 2 relaciona as seqüências que apresentaram mutações na transcriptase reversa que diminuem a susceptibilidade aos NRTI com o ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos NRTI.

**Quadro 2. Descrição das seqüências que apresentaram mutações na transcriptase reversa que diminuem a susceptibilidade aos inibidores de transcriptase reversa análogos de nucleosídeos (NRTI) relacionando-os ao ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos NRTI.**

Seqüência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de transcriptase reversa análogos de nucleosídeos															
				A 62 V	D 67 G	D 67 N	T 69 D	T 69 E	T 69 N	T 69 S	T 69 d	K 70 R	V 75 I	V 75 L	V 75 M	F 77 L	V 118 I	Q 151 M	M 184 V
AG3SP99	1999	SP	B	X															
AJ1SP99	1999	SP	C														X		
AJ2SP99	1999	SP	Recombinante														X		
AM4SP00	2000	SP	B							X									
AP9SP00	2000	SP	B														X		
AQ8SP98	1998	SP	B														X		
AR1SP00	2000	SP	B						X			X							
AR6SP98	1998	SP	B														X		
AS3SP00	2000	SP	B						X										
AX1SP01	2001	SP	B																X
AZ6SP01	2001	SP	B		X		X											X	
BB2SP01	2001	SP	Recombinante	X															
BG3SP98	1998	SP	Recombinante															X	
BG5SP98	1998	SP	B															X	
BH5SP01	2001	SP	B															X	
BH7SP01	2001	SP	B						X										
BI2RS04	2004	RS	B			X									X	X			X
BI3RS04	2004	RS	Recombinante																X
BI4RS04	2004	RS	B			X					X								
BI7SC04	2004	SC	C																X
BK2RS05	2005	RS	F															X	
BL2RS04	2004	RS	Recombinante															X	
BV3RS06	2006	RS	Recombinante													X			

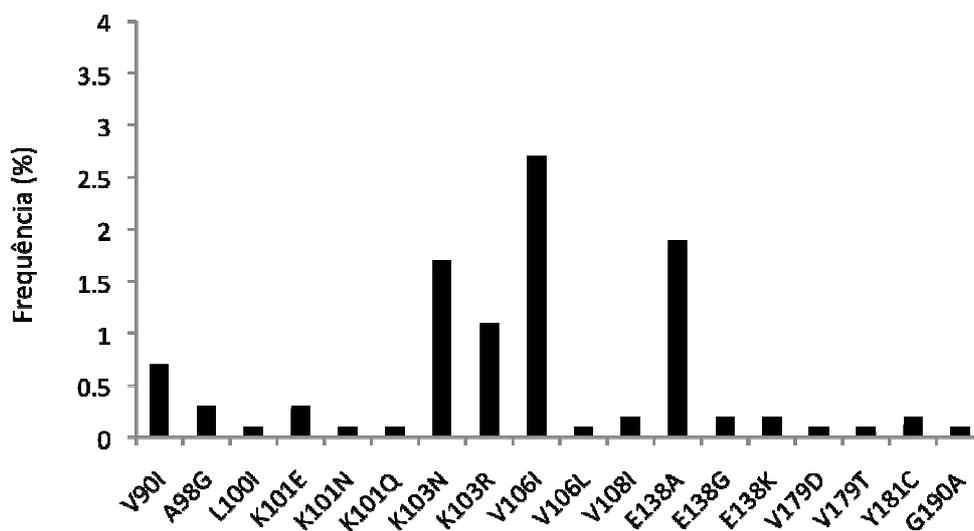


**Quadro 2. Continuação.**

Sequência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de transcriptase reversa análogos de nucleosídeos														
				A 62 V	D 67 G	D 67 N	T 69 D	T 69 E	T 69 N	T 69 S	T 69 d	K 70 R	V 75 I	V 75 L	V 75 M	F 77 L	V 118 I	Q 151 M
DH6SP99	1999	SP	B								X	X						
DM2PR05	2005	PR	B							X								
DM4PR05	2005	PR	C													X		
DN4PR05	2005	PR	B									X						X
DN7PR05	2005	PR	B	X														
DR8SP00	2000	SP	B													X		
DT3SP01	2001	SP	C													X		
DU5SP02	2002	SP	B													X		
DW4SP08	2008	SP	B															X
EB2SP06	2006	SP	B													X		
EB7SP06	2006	SP	B													X		
EC3SP06	2006	SP	B															X
EC5SC05	2005	SC	C															X

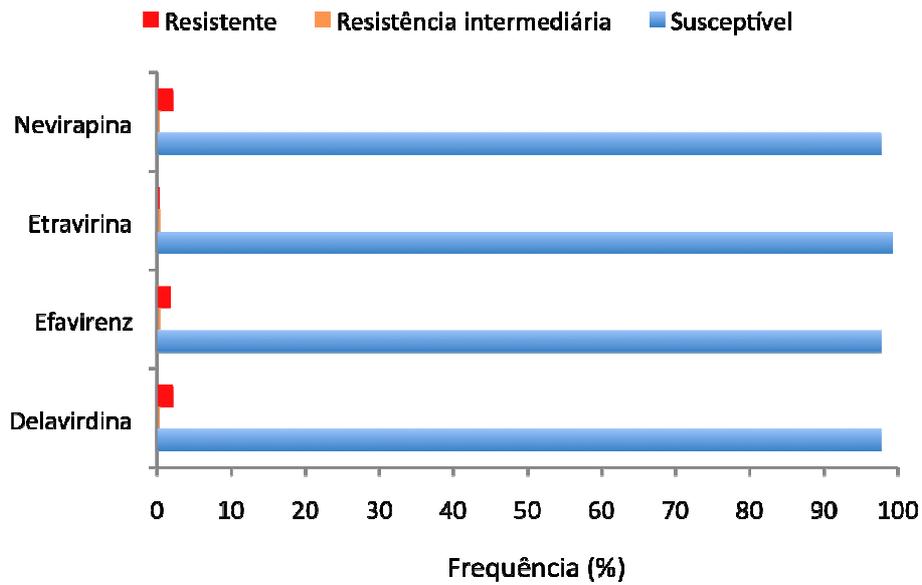
PR, Paraná; RJ, Rio de Janeiro; RS, Rio Grande do Sul; SC, Santa Catarina; e SP, São Paulo.

A análise da transcriptase reversa identificou, ainda, 18 mutações que conferem resistência aos NNRTI: V90I em seis (0,7%) seqüências; A98G e K101E em três (0,3%); L100I, K101N, K101Q, V106L, V179D, V179T e G190A em uma (0,1%); K103N em 15 (1,7%); K103R em 10 (1,1%); V106I em 24 (2,7%); V108I, E138G, E138K e Y181C em duas (0,2%) e E138A em 17 (1,9%) (Figura 17).



**Figura 17. Frequência de mutações associadas à resistência aos inibidores de transcriptase reversa não-análogos de nucleosídeos em isolados de HIV-1 de pacientes não expostos a anti-retrovirais, do Sudeste e Sul do Brasil.**

Concernente a resistência aos NNRTI, 19 (2,1%) amostras foram classificadas como resistentes a DLV e NVP: AO8SP00, AX1SP01, BI2RS04, BI7SC04, BM2RS04, BM5RS04, BP8SC05, BU1RS05, BW3RS04, BZ7SC05, CC1SC05, CO6RJ02, DA1RS04, DA5RS04, DF1RS04, DK2PR05, DO6PR05, DV8SP06 e EC5SC05. Dezesesseis (1,8%) amostras mostraram-se resistentes ao EFV: AX1SP01, BI2RS04, BI7SC04, BM2RS04, BM5RS04, BP8SC05, BU1RS05, BW3RS04, BZ7SC05, CC1SC05, DA1RS04, DF1RS04, DK2PR05, DO6PR05, DV8SP06 e EC5SC05. E três (0,3%), AO8SP00, BI2RS04 e DA1RS04, apresentaram resistência a ETR. As amostras CB9SC05 e DM5PR05 (0,2%) mostraram resistência intermediária a DLV e NVP. Por fim, quatro (0,4%) amostras apresentaram resistência intermediária ao EFV (AO8SP00, CB9SC05, CO6RJ02 e DA5RS04) e a ETR (AX1SP01, CB9SC05, CO6RJ02 e DA5RS04) (Figura 18). Ao todo, 19 (2,1%) e 6 (0,7%) amostras apresentaram resistência e resistência intermediária a pelo menos um NNRTI, respectivamente.



**Figura 18. Níveis de resistência de isolados do HIV-1, de pacientes não expostos a anti-retrovirais, aos inibidores de transcriptase reversa não-análogos de nucleosídeos no Sudeste e Sul do Brasil.**

O Quadro 3 relaciona as seqüências que apresentaram mutações na transcriptase reversa que diminuem a susceptibilidade aos NNRTI com o ano de coleta da amostra, lugar no qual a infecção ocorreu, subtipo do HIV-1 e as respectivas mutações associadas a resistência aos NNRTI.







Quadro 3. Continuação.

Sequência	Ano	Estado	Subtipo	Mutações associadas a resistência aos inibidores de transcriptase reversa não-análogos de nucleosídeos																		
				V 90 I	A 98 G	L 100 I	K 101 E	K 101 N	K 101 Q	K 103 N	K 103 R	V 106 I	V 106 L	V 108 I	E 138 A	E 138 G	E 138 K	V 179 D	V 179 T	Y 181 C	G 190 A	
DJ7PR02	2002	PR	B												X							
DJ8PR02	2002	PR	B												X							
DK2PR05	2005	PR	Recombinante							X												
DM2PR05	2005	PR	B												X							
DM5PR05	2005	PR	B														X					
DM6PR05	2005	PR	B											X								
DN3PR05	2005	PR	Recombinante												X							
DO6PR05	2005	PR	B							X												
DP1PR06	2006	PR	C															X				
DQ5SP99	1999	SP	B									X										
DQ8SP99	1999	SP	B												X							
DR9SP00	2000	SP	Recombinante									X										
DU6SP02	2002	SP	B									X										
DV8SP06	2006	SP	F							X												
DW4SP98	1998	SP	B												X							
DX4SP05	2005	SP	B								X											
DX5SP06	2006	SP	B									X										
EB4SP06	2006	SP	B								X											
EC5SC05	2005	SC	C							X												

PR, Paraná; RJ, Rio de Janeiro; RS, Rio Grande do Sul; SC, Santa Catarina; e SP, São Paulo.

## 4.2. Genotipagem

### 4.2.1. Definição dos Subtipos do HIV-1

Das 895 seqüências analisadas, 587 (65,6%) foram classificadas como subtipo B; 126 (14,1%), como subtipo C; uma (0,1%), como subtipo D; 55 (6,1%), como subtipo F; e 126 (14,1%), como forma recombinante (Figura 19). No entanto, a prevalência dos diferentes subtipos do HIV-1 mostrou-se diferente para cada região analisada (Figura 20 e Apêndice D).

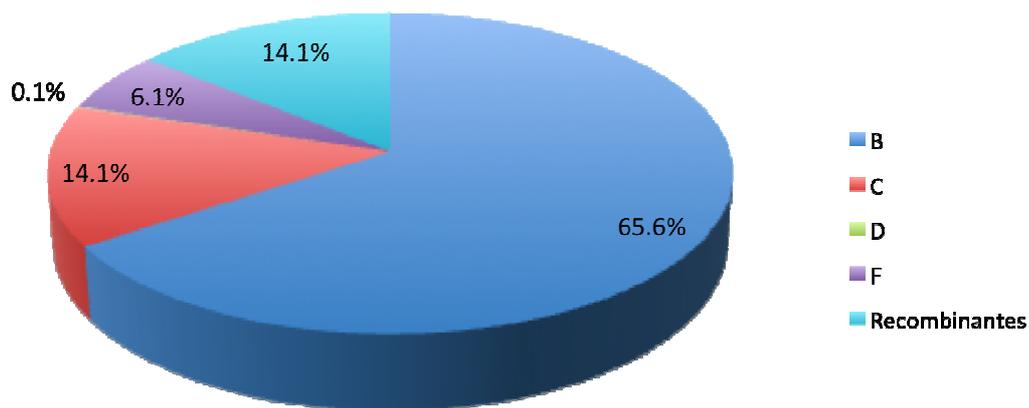


Figura 19. Prevalência dos subtipos e formas recombinantes do HIV-1 isolado de indivíduos não expostos a tratamento anti-retroviral, do Sudeste e Sul do Brasil.

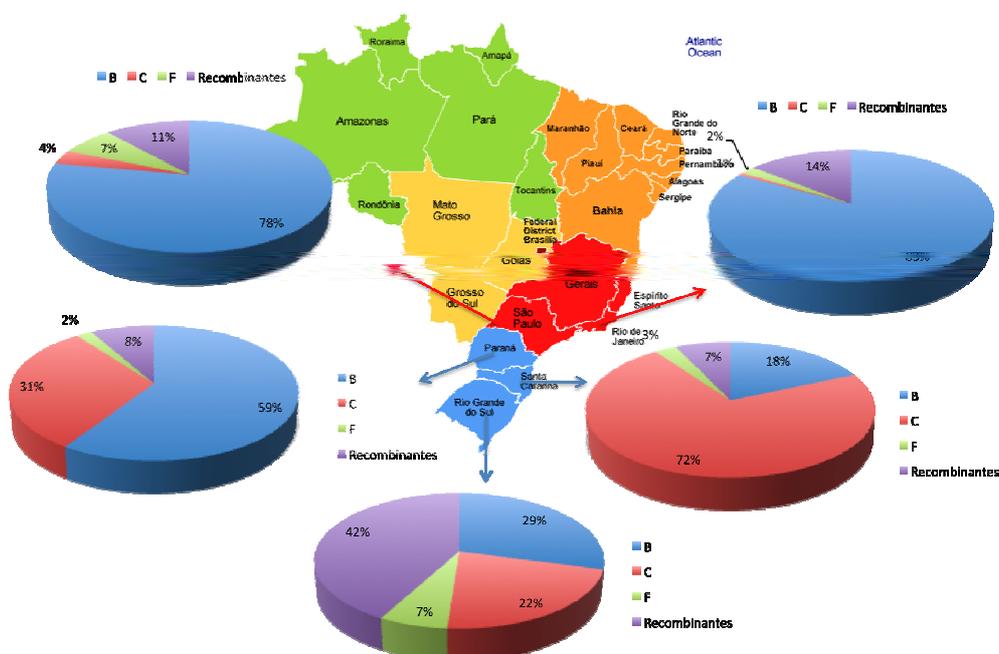


Figura 20. Prevalência dos subtipos e formas recombinantes do HIV-1 isolado de indivíduos não expostos a tratamento anti-retroviral, nos diferentes estados do Sudeste e Sul do Brasil. Uma amostra do subtipo D do Rio de Janeiro não está ilustrada na figura.

No Rio de Janeiro, o subtipo B foi o mais prevalente, ocorrendo em mais de 80% (n=132, 87,4%) das amostras, seguido do subtipo F, presente em 3,3% (n=5). Foi encontrado um (0,7%) caso de subtipo D e um (0,7%) de subtipo C. As formas recombinantes representaram 7,9% (n=12) dos casos no Rio de Janeiro.

Em São Paulo, das 447 amostras analisadas, 360 (80,5%) foram do subtipo B; 16 (3,6%), subtipo C; 33 (7,4%), subtipo F; e 38 (8,5%), foram classificadas como formas recombinantes.

No Paraná, assim como no Sudeste, o subtipo B também foi o mais prevalente, ocorrendo em 59,2% (n=29) das amostras. No entanto, o subtipo C já aparece em 30,6% (n=15) dos casos, havendo apenas um (2,0%) caso de subtipo F e quatro (8,2%) formas recombinantes.

Em Santa Catarina, o subtipo C ocorreu em 69,4% (n=52) das seqüências, sendo seguido pelo subtipo B, cuja prevalência foi de 21,3% (n=16). O subtipo F e as formas recombinantes foram detectadas em apenas três (4,0%) e quatro (5,3%) amostras, respectivamente.

No Rio Grande do Sul, a prevalência dos subtipos B e C foram muito similares, 28,9% (n=50) e 24,3% (n=42), respectivamente. O subtipo F apareceu em 7,5% (n=13) e as formas recombinantes em 39,3% (n=68), sendo, portanto, o estado no qual foi identificado o maior número de recombinantes.

#### 4.2.2. Definição das Formas Recombinantes do HIV-1

Das 126 seqüências classificadas como recombinantes pelas análises filogenéticas e de recombinação, 87 tiveram o perfil de recombinação definido pelo programa RDP3 (Martin, Williamson *et al.*, 2005). As seqüências de cada perfil recombinante foram agrupadas de acordo com a posição dos pontos de recombinação. Essa, por sua vez, variou em até 20 nucleotídeos dentro de cada grupo definido, sendo descrita a numeração daquela correspondente à primeira amostra relatada do grupo em questão, de acordo com o genoma de HXB2 (Figura 21).

Sessenta e três seqüências apresentaram recombinação entre os subtipos B e C do HIV-1. Dessas, 11 exibiram o perfil de recombinação PR<sup>B/C</sup>/RT<sup>C/B</sup>; 47, PR<sup>C</sup>/RT<sup>C/B/C</sup>; e cinco, PR<sup>C/B</sup>/RT<sup>B/C</sup>.

Das formas recombinantes PR<sup>B/C</sup>/RT<sup>C/B</sup>, as seqüências AA9SP99 e DH1RS04 apresentaram os pontos de recombinação nas posições 2321 e 2974; BI3RS04, BL2RS04 e BL6RS04, nas posições 2320 e 2949; BI6SC04, nas posições 2361 e

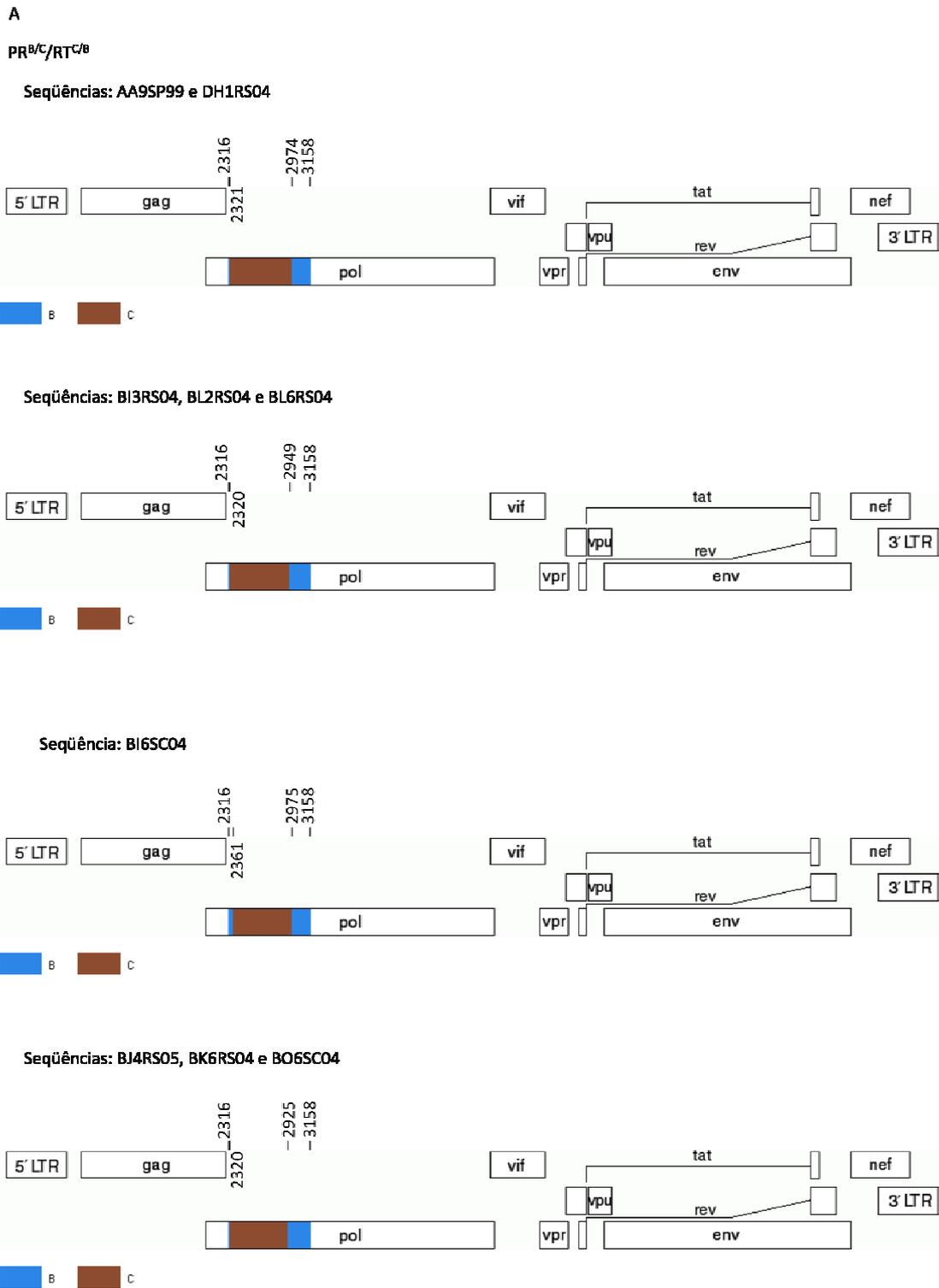
2975; BJ4RS05, BK6RS04 e BO6SC04, nas posições 2320 e 2925; e BN2RS05 e CX1RS04, nas posições 2320 e 2910.

Concernente ao perfil  $PR^C/RT^{C/B/C}$ , nas amostras AC5SP99 e BT2RS05, os pontos de recombinação ocorreram nos nucleotídeos 2701 e 3100; em BM7RS04, nos nucleotídeos 2926 e 3103; em BP8SC05, BQ1SC05, BQ2RS04, BR3RS05, BS8RS05, BU1RS05, BU6RS05, BU7RS05, BV1RS05, BV3RS06, BV4RS05, BV5RS05, BW3RS04, BW5RS04 e BW9RS04, nos nucleotídeos 2909 e 3135; em BQ3RS05, BR8RS05, CX7RS04, CY4RS04, CZ6RS04, DD7RS04, DE5RS04 e DF9RS04, nos nucleotídeos 2950 e 3135; em BQ4RS05, BR2RS06, BS1RS06, DA3RS04, DD4RS04 e DF8RS04, nos nucleotídeos 2970 e 3135; em BQ7RS04, nos nucleotídeos 2904 e 3109; em BR7RS04, BT9RS05, CX8RS04, DB1RS04, DD6RS04, DE6RS04, DF3RS04 e DF6RS04, nos nucleotídeos 2966 e 3148; em BT8RS05, nos nucleotídeos 2934 e 3148; em CX2RS04, nos nucleotídeos 2981 e 3154; e em CZ2RS04, CZ9RS04, DD3RS04 e DG3RS04, nos nucleotídeos 2946 e 3148.

Em relação às formas  $PR^{C/B}/RT^{B/C}$ , em AC8SP99, os pontos de recombinação ocorreram nas posições 2506 e 3092; em BM8RS04, nas posições 2495 e 3124; em DM1PR05, nas posições 2320 e 2977; em DM3PR05, nas posições 2320 e 2960; e em DN3PR05, nas posições 2320 e 2947.

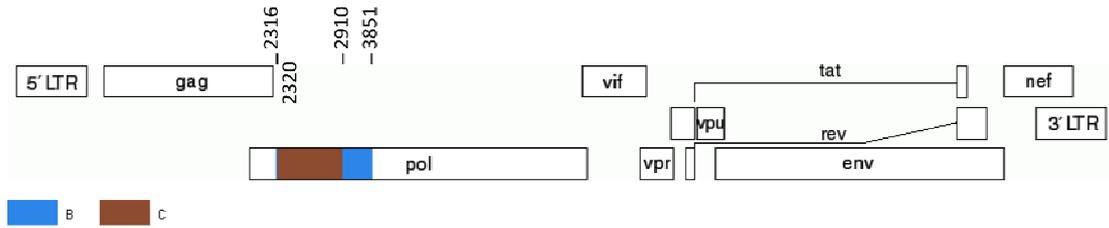
A recombinação entre os subtipos B e F foi descrita em 22 amostras. Dessas, três apresentaram o perfil  $PR^{B/F}/RT^{F/B}$ ; 11,  $PR^F/RT^{F/B/F}$ ; e oito,  $PR^{F/B}/RT^{B/F}$ . Das formas  $PR^{B/F}/RT^{F/B}$ , a recombinação em AC6SP98 ocorreu nas posições 2320 e 2729; em AD4SP99, nas posições 2361 e 2719; e em BG3SP98, nas posições 2352 e 2933. Quanto aos perfis  $PR^F/RT^{F/B/F}$ , as amostras AJ2SP99, AJ3SP98, AN9SP00, AP6SP00, DR9SP00 e DS9SP00 apresentaram pontos de recombinação nos nucleotídeos 2758 e 3118; AM5SP98, nos nucleotídeos 2758 e 3143; AR7SP00, nos nucleotídeos 2785 e 3119; AX7SP01, nos nucleotídeos 2839 e 3119; AZ2SP01, nos nucleotídeos 2455 e 3096; BA2SP01, nos nucleotídeos 2693 e 3096; e DH3SP99, nos nucleotídeos 2715 e 3121. Dos perfis  $PR^{F/B}/RT^{B/F}$ , as seqüências BA7SP01, BB2SP01 e BE3SP02 recombinaram nas posições 2749 e 3046; DI1SP99, DI2SP00 e DI3SP01, nas posições 2485 e 3132; e DV5SP04, nas posições 2476 e 3151.

Duas amostras apresentaram recombinação entre os subtipos F e C. Em BM3RS04 ( $PR^C/RT^{C/F/C}$ ), a passagem do subtipo C para F ocorreu na posição 2852 e de F para, na 3132. DD2RS04 ( $PR^{F/C}/RT^{C/F}$ ), apresentou os pontos de recombinação nos nucleotídeos 2488 e 3120.



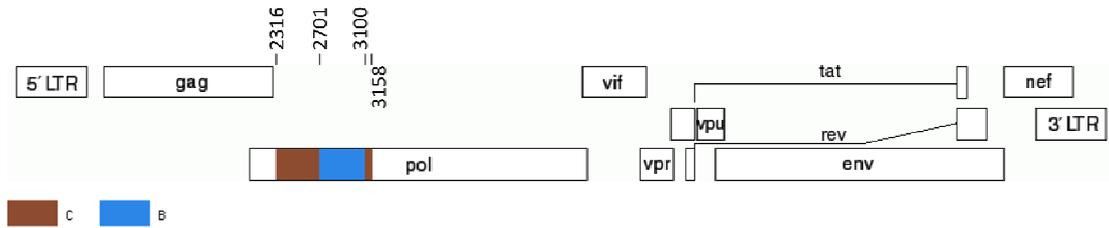
**Figura 21. Representação esquemática das formas recombinantes do HIV-1 descritas no Sudeste e Sul do Brasil, baseada em *pol*, de acordo com o caracterizado pelo programa *Recombination Detection Program* – RDP (Martin, Williamson *et al.*, 2005). Foram identificadas recombinações entre os subtipos do HIV-1 B e C (A); B e F (B); e C e F (C).**

Seqüências: BN2RS05 e CX1RS04

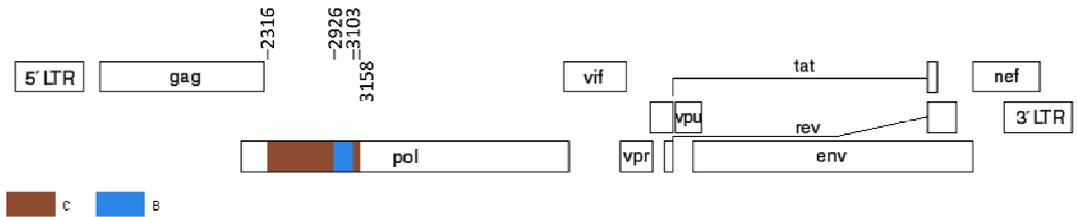


PR<sup>C</sup>/RT<sup>C</sup>/B<sup>C</sup>

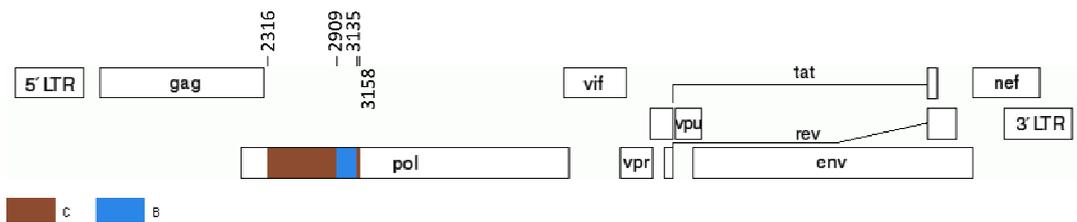
Seqüências: AC5SP99 e BT2RS05



Seqüência: BM7RS04



Seqüências: BP8SC05, BQ1SC05, BQ2RS04, BR3RS05, BS8RS05, BU1RS05, BU6RS05, BU7RS05, BV1RS05, BV3RS06, BV4RS05, BV5RS05, BW3RS04, BW3RS04, BW5RS04 e BW9RS04



Seqüências: BQ3RS05, BR8RS05, CX7RS04, CY4RS04, CZ6RS04, DD7RS04, DE5RS04 e DF9RS04

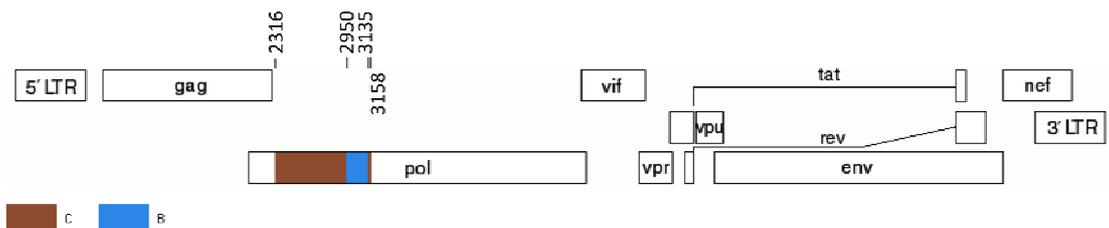
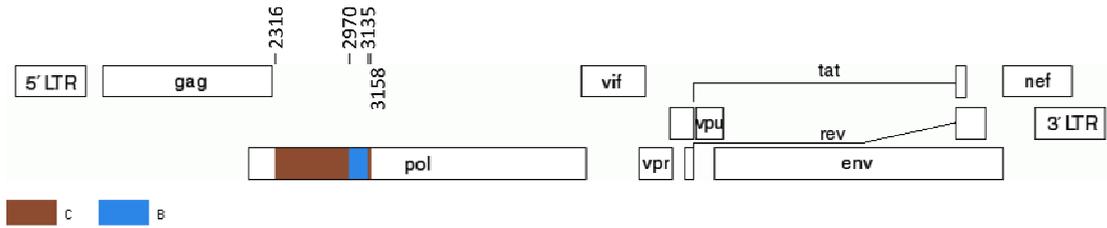
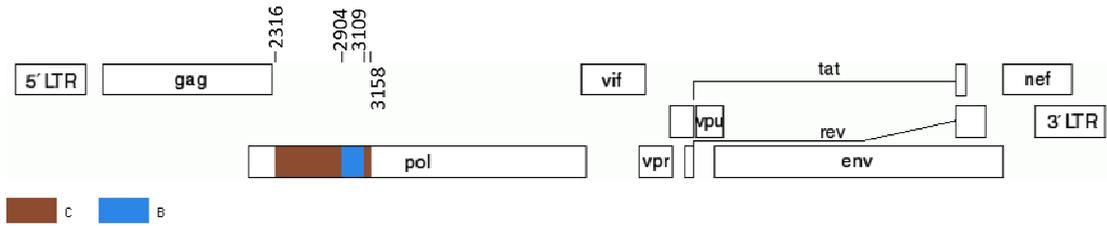


Figura 21. Continuação.

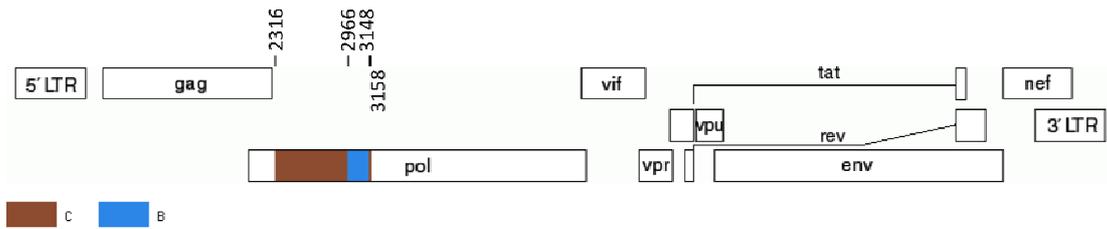
Seqüências: BQ4RS05, BR2RS06, BS1RS06, DA3RS04, DD4RS04 e DF8RS04



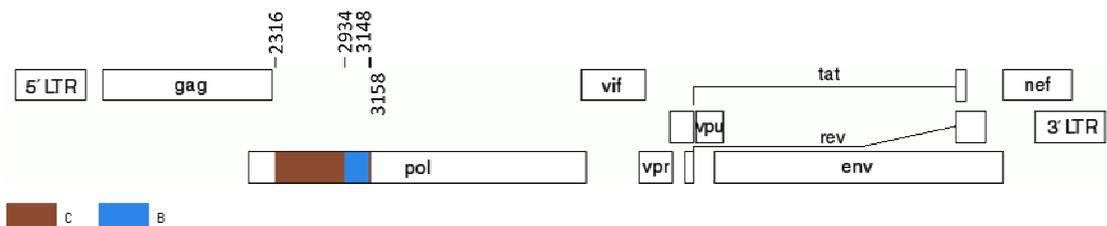
Seqüência: BQ7RS04



Seqüências: BR7RS04, BT9RS05, CX8RS04, DB1RS04, DD6RS04, DE6RS04, DF3RS04, DF6RS04



Seqüência: BT8RS05



Seqüência: CX2RS04

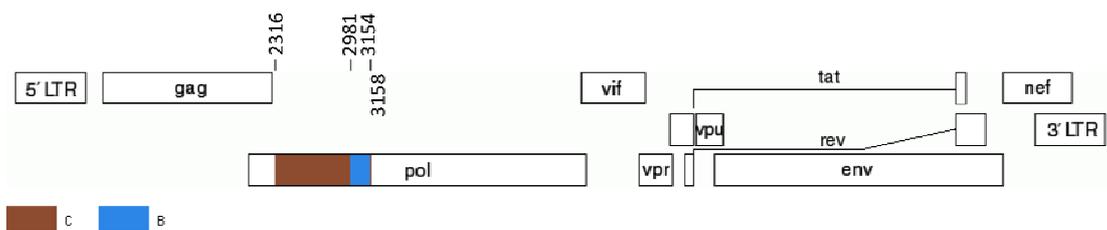
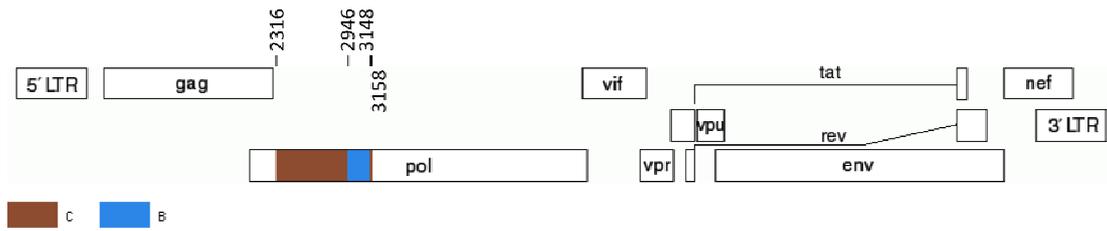
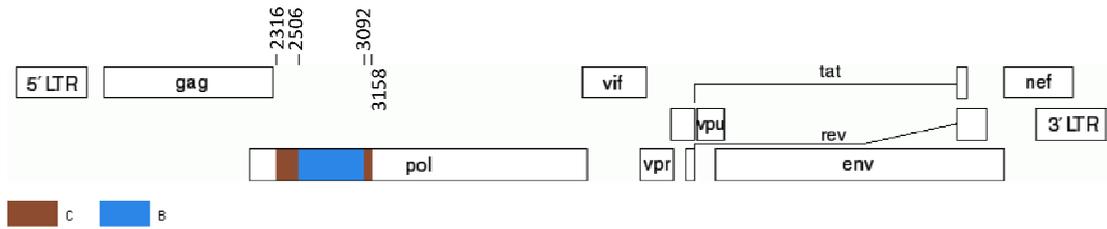


Figura 21. Continuação

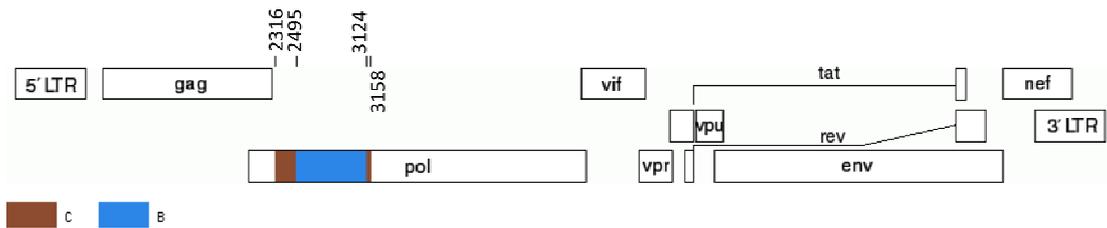
Seqüência: CZ2RS04

PR<sup>C</sup>/RT<sup>C</sup>/B<sup>C</sup>

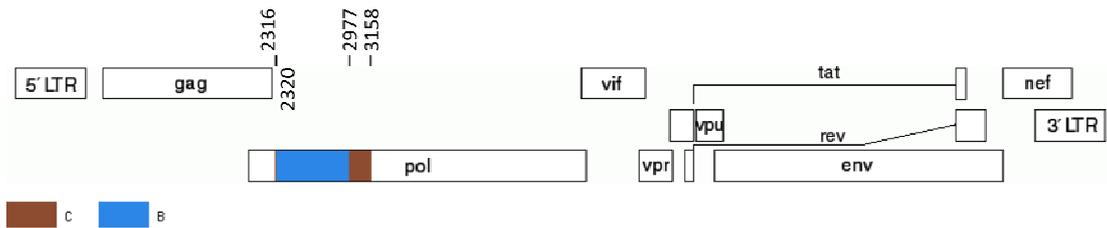
Seqüência: AC8SP99



Seqüência: BM8RS04



Seqüência: DM1PR05



Seqüência: DM3PR05

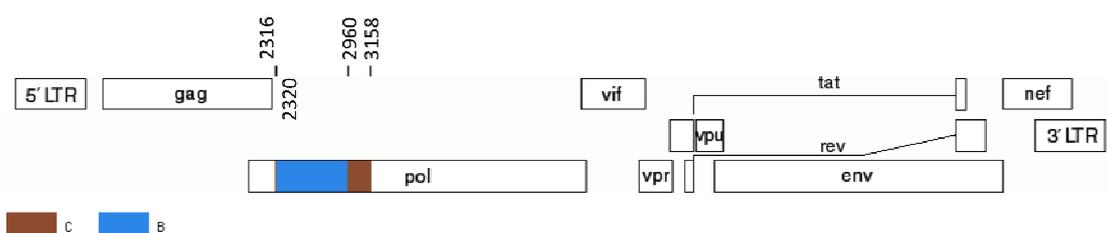
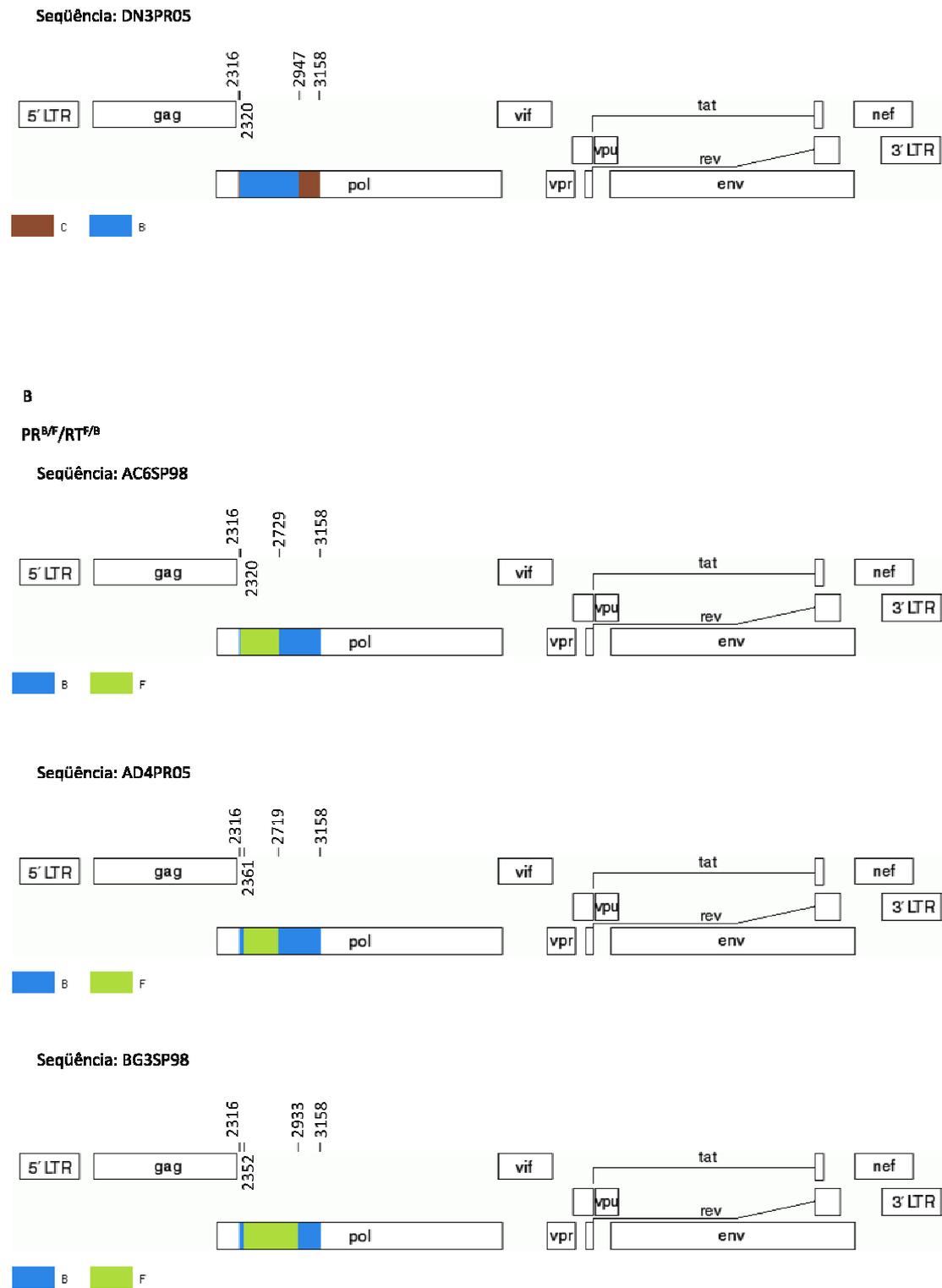


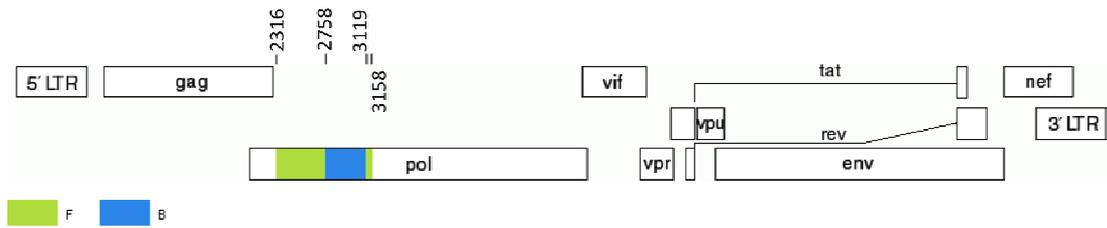
Figura 21. Continuação



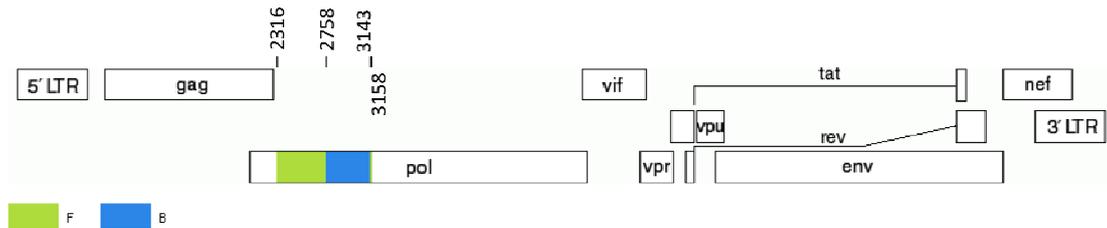
**Figura 21. Continuação**

PR<sup>F</sup>/RT<sup>F/B/F</sup>

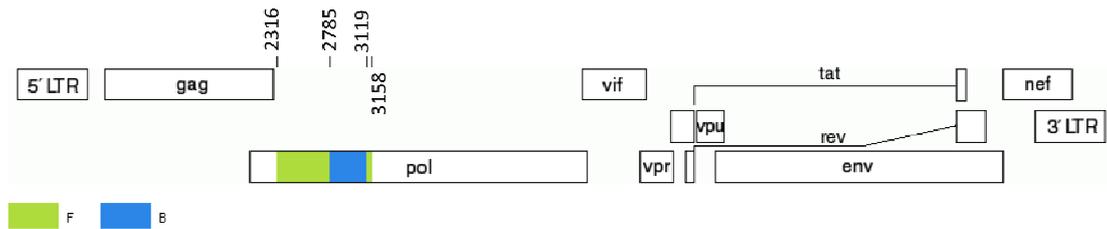
Seqüências: AJ2SP99, AJ3SP98, AN9SP00, AP6SP00, DR9SP00, DS9SP00



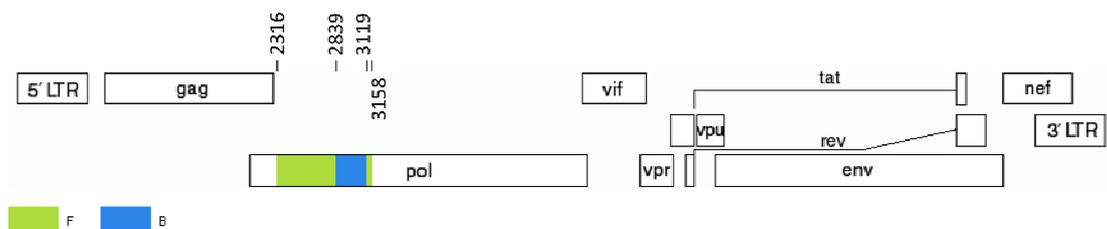
Seqüência: AM5SP98



Seqüência: AR7SP00



Seqüência: AX7SP01



Seqüência: AZ2SP01

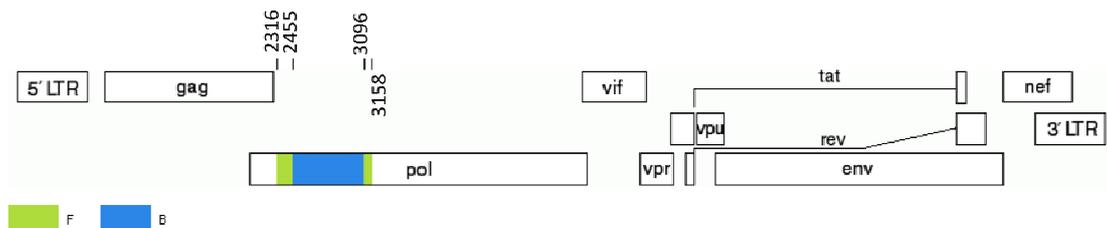
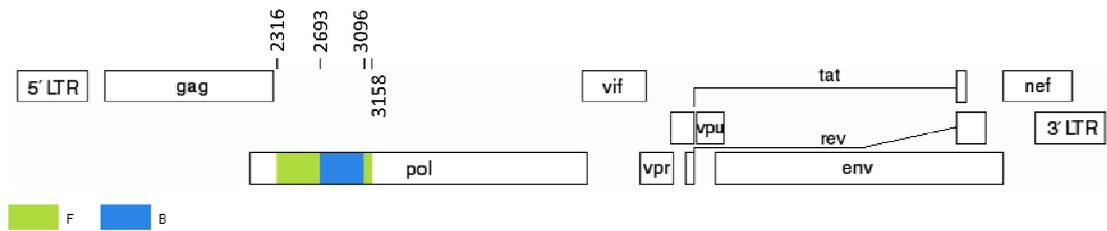
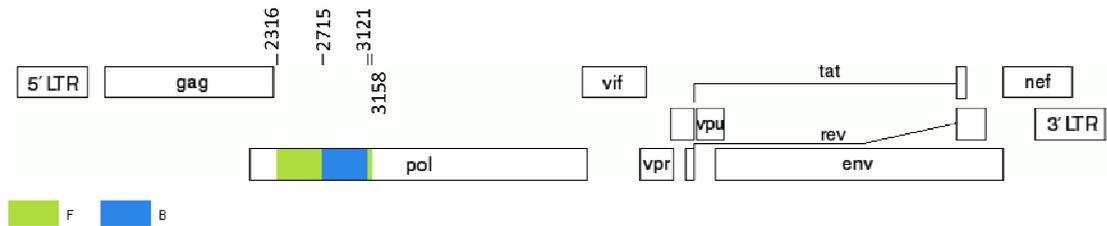


Figura 21. Continuação

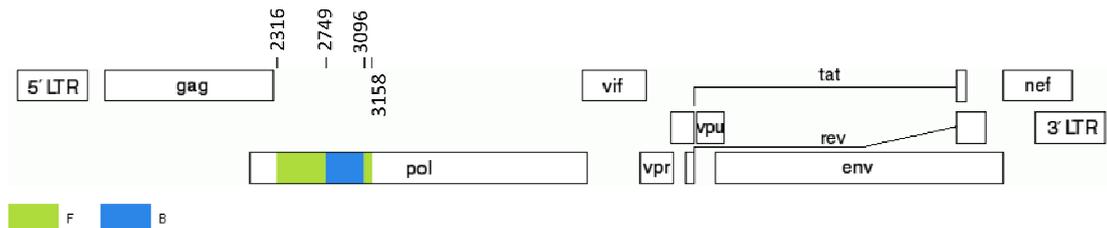
## Seqüência: BA2SP01



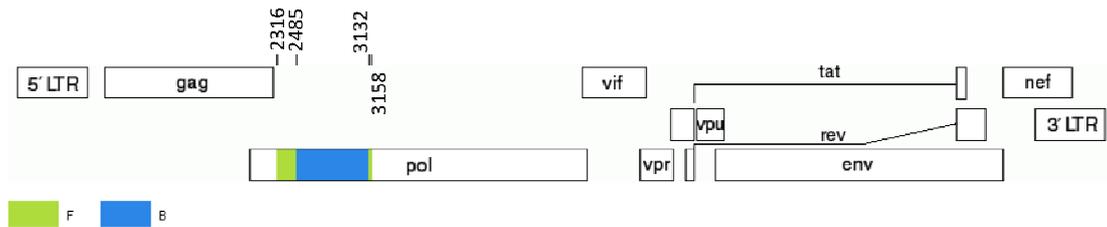
## Seqüência: DH3SP99

PR<sup>F</sup>/B/RT<sup>B</sup>/F

## Seqüência: BA7SP01, BB2SP01 e BE3SP02



## Seqüência: DI1SP99, DI2SP00 e DI3SP01



## Seqüência: DV5SP04

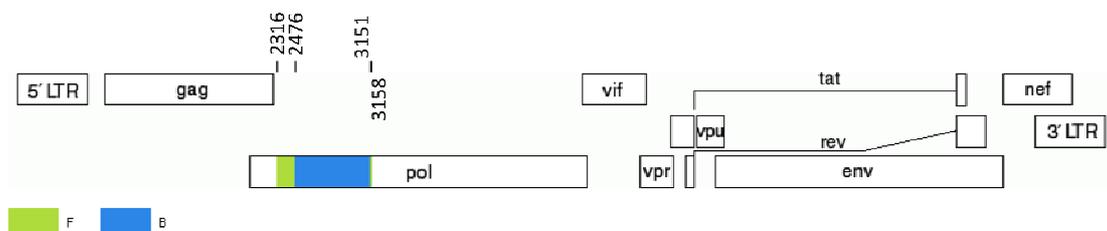
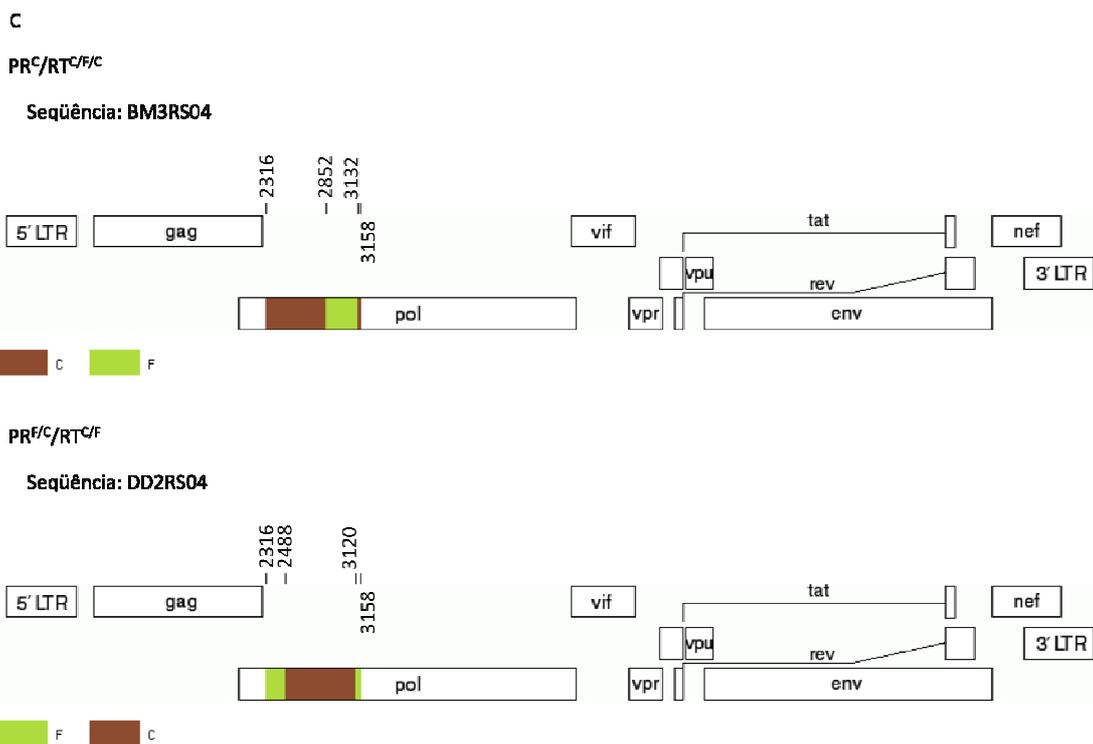


Figura 21. Continuação



**Figura 21. Continuação**

### 4.3. Análise da Presença de Assinaturas na Protease e Transcriptase Reversa

No subtipo B, a mutação I135T, na transcriptase reversa, foi encontrada em 58,3% dos isolados analisados para as regiões Sudeste e Sul.

Nas seqüências do subtipo C, as prováveis assinaturas N37K e R41N, na protease, tiveram a prevalência de 88,9% e 90,5%, respectivamente.

Por sua vez, o subtipo F apresentou duas mutações prevalentes em mais da metade das amostras: Q61N (63,6%) na protease e E169D (78,2%) transcriptase reversa.

A Tabela 2 descreve a prevalência das assinaturas encontradas da PR e RT nas seqüências do subtipo B, C e F circulantes no Brasil.

**Tabela 2. Prevalência das assinaturas encontradas da protease (PR) e transcriptase reversa (RT) nas seqüências do subtipo B, C e F circulantes no Sudeste e Sul do Brasil.**

Subtipo	Gene	Mutação	Prevalência
B	RT	I135T	58,3%
C	PR	N37K	88,9%
C	PR	R41N	90,5%
F	PR	Q61N	63,6%
F	RT	E169D	78,2%

Das 85 seqüências que recombinaram com o subtipo B, a substituição I135T foi encontrada em 40 (47,1%) amostras. Das 65 formas recombinantes descritas envolvendo o subtipo C, 58 (89,2%) apresentaram a mutação N37K e 50 (76,9%), a R41N, ambas na protease. E por fim, as mutações Q61N e E169D foram descritas em 19 (79,2%) e quatro (16,7%), das 24 formas recombinantes com similaridade ao subtipo F (Tabela 3).

**Tabela 3. Identificação das assinaturas encontradas da protease (PR) e transcriptase reversa (RT) nas seqüências do subtipo B, C e F circulantes no Sudeste e Sul do Brasil, nas formas recombinantes caracterizadas por análise filogenética de máxima verossimilhança e pelo programa *Recombination Detection Program* – RDP (Martin, Williamson *et al.*, 2005).**

Seqüência	Protease			Transcriptase reversa	
	N	R	Q	I	E
	37	41	61	135	169
	K	N	N	T	D
AA7SP99	N	K	D	I	E
<b>AA9SP99</b>	K	N	Q	R	E
AC5SP99	S	K	D	I	X
AC6SP98	N	K	N	I	E
AC8SP99	N	K	Q	I	E
AD4SP99	N	K	N	T	E
AF3SP99	N	K	N	I	E
AJ2SP99	N	K	N	I	E
AJ3SP98	N	K	N	I	E
AM5SP98	S	R	N	V	E
AN9SP00	N	K	N	T	E
AP6SP00	N	K	D	I	E
AR7SP00	D	K	N	T	E
AS5SP00	N	K	Q	V	E

Tabela 3. Continuação

Seqüência	Protease			Transcriptase reversa	
	N 37 K	R 41 N	Q 61 N	I 135 T	E 169 D
<b>AX7SP01</b>	N	K	N	I	E
AY2SP01	S	K	Q	I	X
<b>AZ2SP01</b>	N	K	N	T	E
AZ5SP01	N	K	N	T	E
<b>BA2SP01</b>	N	K	D	T	E
BA7SP01	N	K	D	X	E
BB2SP01	N	K	N	L	D
BB6SP98	N	K	N	V	E
<b>BE3SP02</b>	E	R	N	T	E
BF1SP02	N	K	Q	T	X
<b>BG3SP98</b>	N	K	N	I	E
<b>BI3PA04</b>	K	N	Q	T	E
<b>BI6PA04</b>	R	N	Q	I	E
<b>BJ4PA05HE</b>	K	N	Q	I	E
<b>BK6PA04HE</b>	K	D	Q	T	E
<b>BL2PA04HE</b>	K	N	Q	K	E
BL3PA04	N	K	N	T	D
<b>BL6PA04</b>	K	N	Q	T	E
BL9PA04HE	N	K	N	L	E
<b>BM3PA04HE</b>	K	I	Q	I	D
<b>BM7PA04HE</b>	K	N	Q	T	E
<b>BM8PA04HE</b>	K	G	Q	T	E
<b>BN2PA05HE</b>	K	N	Q	I	E
<b>BO6SC04</b>	K	N	Q	T	E
<b>BP8SC05</b>	K	N	Q	I	E
<b>BQ1SC05</b>	K	N	Q	T	E
<b>BQ2PA04HE</b>	K	N	Q	T	E
<b>BQ3PA05HE</b>	N	K	H	I	E
<b>BQ4PA05HE</b>	K	N	Q	I	D
<b>BQ7PA04HE</b>	K	N	Q	T	E
<b>BR2PA06</b>	K	N	Q	T	E
<b>BR3PA05HE</b>	K	N	Q	T	E
<b>BR7PA04HE</b>	K	N	Q	I	E
<b>BR8PA05HE</b>	K	I	Q	I	E
<b>BS1PA06</b>	K	N	Q	I	E
BS6PA05	A	K	Q	I	E
<b>BS8PA05HE</b>	R	N	Q	T	E
BS9PA05HE	N	K	Q	I	E
<b>BT2PA05HE</b>	K	N	Q	T	E
BT3PA05HE	K	N	E	I	E
BT5PA05HE	D	K	Q	V	E
<b>BT8PA05HE</b>	N	K	N	I	E
<b>BT9PA05HE</b>	K	N	Q	T	E

Tabela 3. Continuação

Seqüência	Protease			Transcriptase reversa	
	N	R	Q	I	E
	37	41	61	135	169
	K	N	N	T	D
<b>BU1PA05HE</b>	K	N	Q	T	E
<b>BU6PA05BI</b>	K	I	Q	I	E
<b>BU7PA05HO</b>	K	I	Q	I	E
BU8PA05HE	K	S	Q	I	E
<b>BV1PA05HE</b>	K	N	N	T	E
<b>BV3PA06</b>	K	N	Q	I	E
<b>BV4PA05BT</b>	K	N	Q	T	E
<b>BV5PA05HE</b>	K	N	Q	I	E
BV9PA05HE	N	R	Q	I	E
<b>BW3PA04HE</b>	K	P	Q	T	E
<b>BW5PA04HE</b>	K	N	Q	I	E
<b>BW9PA04HE</b>	K	N	Q	I	E
CI7RJ98	K	N	Q	I	E
CO4RJ02VT	N	K	N	I	E
<b>CP1RJ02VT</b>	N	K	Q	T	D
CQ4RJ03VT	N	K	Q	S	E
CQ6RJ03VT	N	K	D	V	E
CU9RJ02	N	K	N	L	D
CV1RJ02	S	K	Q	V	E
CV3RJ02	N	K	D	I	E
<b>CV4RJ02</b>	N	R	Q	I	D
CV5RJ02	N	K	N	V	D
<b>CV6RJ02</b>	N	R	Q	V	D
CV8RJ02	N	K	D	I	D
CW5PA04	K	N	Q	I	E
<b>CX1PA04</b>	K	N	Q	I	E
<b>CX2PA04</b>	K	N	Q	T	E
CX3PA04	K	N	Q	T	E
<b>CX7PA04</b>	K	N	Q	I	E
<b>CX8PA04</b>	K	N	Q	I	E
<b>CY4PA04</b>	K	N	Q	T	E
<b>CZ2PA04</b>	K	N	Q	T	E
<b>CZ6PA04</b>	K	N	Q	I	E
<b>CZ9PA04</b>	K	N	Q	T	E
DA2PA04	N	K	N	T	E
<b>DA3PA04</b>	K	N	Q	T	N
<b>DB1PA04</b>	K	I	Q	T	E
<b>DD2PA04</b>	N	K	N	T	E
<b>DD3PA04</b>	K	D	Q	T	E
<b>DD4PA04</b>	K	N	Q	I	E
<b>DD6PA04</b>	K	N	Q	I	E
<b>DD7PA04</b>	K	N	Q	T	E

Tabela 3. Continuação

Seqüência	Protease			Transcriptase reversa	
	N 37 K	R 41 N	Q 61 N	I 135 T	E 169 D
DE2PA04	K	E	N	I	E
<b>DE5PA04</b>	K	N	Q	V	E
<b>DE6PA04</b>	K	N	Q	T	E
<b>DF3PA04</b>	K	N	Q	T	E
<b>DF6PA04</b>	K	N	Q	I	E
<b>DF8PA04</b>	K	N	Q	I	E
<b>DF9PA04</b>	K	N	Q	I	E
DG1PA04	D	K	E	T	E
<b>DG3PA04</b>	K	N	Q	T	E
<b>DH1PA04</b>	K	A	Q	T	E
<b>DH3SA99HE</b>	N	K	N	M	E
<b>DI1SA99HE</b>	N	K	N	I	E
<b>DI2SA00HE</b>	N	K	N	T	D
<b>DI3SA01</b>	N	K	D	V	E
DI5SP03HO	S	K	Q	T	E
DK2CT05	K	N	Q	R	E
<b>DM1CT05HE</b>	K	N	Q	I	E
<b>DM3CT05HE</b>	K	N	Q	I	E
<b>DN3CT05HE</b>	K	N	Q	T	E
DQ9SP99	N	K	N	T	D
<b>DR9SP00</b>	E	K	N	X	E
<b>DS9SP00</b>	N	K	N	T	D
DU8SP03	S	R	E	I	E
<b>DV5SP04</b>	N	K	N	T	E
DW9SP07	N	K	N	I	E
DX7RP04	N	K	Q	I	E
DY5RP05	N	R	Q	I	E

As assinaturas identificadas nas formas recombinantes aparecem destacadas em cinza.  
 As formas recombinantes caracterizadas pelo programa RDP (Martin, Williamson *et al.*, 2005) estão destacadas em negrito.

#### **4.4. Identificação de Epitopos na Protease e Transcriptase Reversa**

A presença de epitopos reconhecidos por linfócitos T CD4<sup>+</sup> e T CD8<sup>+</sup> na protease e transcriptase reversa foi verificada nas 895 sequencias selecionadas por esse estudo, baseando-se no *Los Alamos HIV Immunology database* (Los Alamos, 2009a), de acordo com o descrito na metodologia.

Foram identificados cinco epitopos reconhecidos por linfócitos T CD8<sup>+</sup>, na protease, e 15 na RT, sendo 13 epitopos reconhecidos por linfócitos T CD8<sup>+</sup> e dois por T CD4<sup>+</sup>.

A Tabela 4 relaciona cada epitopo encontrado ao componente do sistema immune pelo qual é reconhecido, indica a posição do epitopo em relação a HXB2, a linhagem de HIV-1 que foi primeiramente identificado e a frequência com a qual foi encontrada no Brasil em pacientes infectados por HIV e que não receberam tratamento anti-retroviral, em cada um dos subtipos encontrados e no total de amostras analisadas.

**Tabela 4. Caracterização dos epitopos descritos na protease e transcriptase reversa do HIV-1 circulante no Sudeste e Sul do Brasil.** A porcentagem reflete a prevalência do epítipo em cada um dos subtipos encontrados e no total de amostras analisadas.

Epítipo	Componente do sistema imune	Posição em relação a HXB2	Subtipos de HIV-1 em que foi descrito	Frequência (%) no Brasil				
				Subtipo B	Subtipo C	Subtipo F	Recombinantes	Total
EEMNLPGRW	Linfócitos T CD8 <sup>+</sup>	PR - aas 34-42	Subtipo B	23,2	0,0	0,0	3,2	15,6
RQYDQILIEI	Linfócitos T CD8 <sup>+</sup>	PR - aas 57-66	Subtipo B	12,3	48,4	3,6	30,2	19,3
GKKAIGTVL	Linfócitos T CD8 <sup>+</sup>	PR - aas 68-76	Subtipos B e C	1,9	87,3	0,0	47,6	20,2
KAIGTVLV	Linfócitos T CD8 <sup>+</sup>	PR - aas 70-77	Subtipos B e C	56,9	87,3	49,1	67,5	62,1
LVGPTPVNI	Linfócitos T CD8 <sup>+</sup>	PR - aas 76-84	Subtipos B e C	75,8	93,6	74,5	87,3	79,8
KLVDFRELNK	Linfócitos T CD8 <sup>+</sup>	RT - aas 37-82	Subtipo B	93,7	96,8	96,4	95,2	94,4
GIPHPAGLK <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 93-101	Subtipos B, C, CRF01_AE	87,0	96,0	92,7	94,4	89,7
TVLVDVGDAY	Linfócitos T CD8 <sup>+</sup>	RT - aas 107-115	Subtipos B, C, A/G	95,7	99,2	96,4	98,4	96,5
VPLDEDFRKY	Linfócitos T CD8 <sup>+</sup>	RT - aas 118-127	Subtipos B e C	19,4	30,0	9,1	15,9	19,3
YTAFTIPSV	Linfócitos T CD8 <sup>+</sup>	RT - aas 127-135		1,7	18,2	41,8	7,9	5,2
TAFTIPSI <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 128-135	Subtipos B e C	34,4	55,6	0,0	43,6	28,8
NETPGIRYQY <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 137-146	Subtipos B e C	78,4	92,1	87,3	89,7	82,5
IRYQYNVL <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 142-149	Subtipo C	81,8	94,4	87,3	91,3	85,2
SPAIFQSSM <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 156-164	Subtipos A1, B, C, D, A/E	76,8	50,8	5,4	75,4	68,6
AIFQSSMTK <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 158-166	Subtipos A, B, C, D, CRF01_AE	65,2	39,7	3,6	65,1	57,9
KQNPDIYIY	Linfócitos T CD8 <sup>+</sup>	RT - aas 173-181	Subtipo B	48,4	0,8	0,0	50,0	38,9
HPDIVIYQY	Linfócitos T CD8 <sup>+</sup>	RT - aas 175-183	Subtipos A e B	0,5	0,0	0,0	0,0	0,3
NPDEVIYQY	Linfócitos T CD8 <sup>+</sup>	RT - aas 175-183	Subtipo B	19,9	65,9	9,1	18,2	25,6
VIYQYMDL <sup>*</sup>	Linfócitos T CD8 <sup>+</sup>	RT - aas 179-187	Subtipo C	91,3	90,5	92,7	92,9	91,5
FRKYTAFTIPSINNE <sup>*</sup>	Linfócitos T CD4 <sup>+</sup>	RT - aas 124-138		32,4	54,0	30,9	43,6	37,0
SPAIFQSSMTKILEP <sup>*</sup>	Linfócitos T CD4 <sup>+</sup>	RT - aas 156-170		59,8	39,7	0,0	60,3	53,4

<sup>a</sup> em relação a linhagem de referência do HIV-1 HXB2.

<sup>\*</sup> epítipos também descritos na amostra de subtipo D (CD9RJ96).

## **4.5. Análise Filogenética dos Subtipos B e C Circulantes no Brasil**

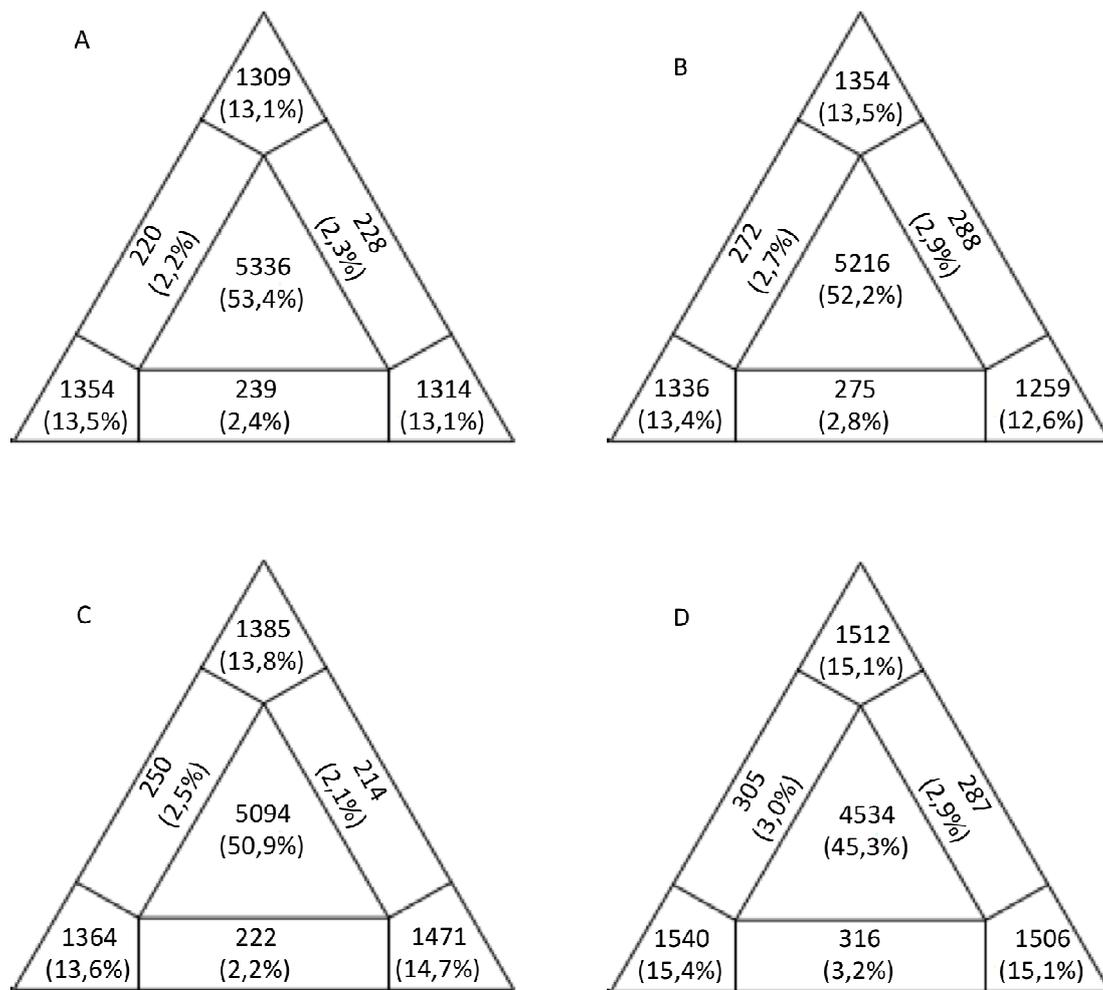
### **4.5.1. Subtipo B**

#### **4.5.1.1. Mapeamento de Verossimilhança**

O mapeamento de verossimilhança de 10.000 quartetos randômicos foi realizado para os alinhamentos dos genes da RT e gp120, incluindo apenas as seqüências brasileiras e ambas brasileiras e globais (Figura 22).

Nas análises da RT e gp120, considerando tanto seqüências brasileiras quanto globais, mais de 52% dos quartetos foram localizados no centro do mapa, indicando sinal filogenético do tipo estrela elevado para ambos os alinhamentos.

Da mesma forma, as análises incluindo apenas seqüências brasileiras também revelaram sinal filogenético do tipo estrela significativo tanto para o alinhamento da RT (50,9%) quanto para o de gp120 (45,3%). Essa elevada proporção de quartetos localizados no centro do triângulo está de acordo com a dispersão extensiva do subtipo B no Brasil, representando até 90% das linhagens de HIV-1 circulantes no Nordeste (Couto-Fernandez, Morgado *et al.*, 1999; Gadelha, Shindo *et al.*, 2003) e Centro-Oeste do país (Cerqueira, Amorim *et al.*, 2004; Stefani, Pereira *et al.*, 2007; Veras, Veras *et al.*, 2007).



**Figura 22. Mapeamento de verossimilhança de linhagens do subtipo B do HIV-1 com distribuição global (triângulos superiores) e restrita ao Brasil (triângulos inferiores). As análises foram baseadas no gene da transcriptase reversa (A, 499 seqüências; e C, 317 seqüências) e *gp120* (*env*) (B, 333 seqüências; e D, 150 seqüências). Os números e porcentagens representam a distribuição detalhada dos quartetos em cada região do mapa. Sete regiões principais podem ser distinguidas: os três vértices, representando sinal filogenético típico de árvore (topologia resolvida); os três lados, representando sinal filogenético do tipo rede (com duas possibilidades de topologia); e o centro, representando sinal filogenético do estrela (topologia não resolvida).**

#### 4.5.1.2. Origem do Subtipo B no Brasil

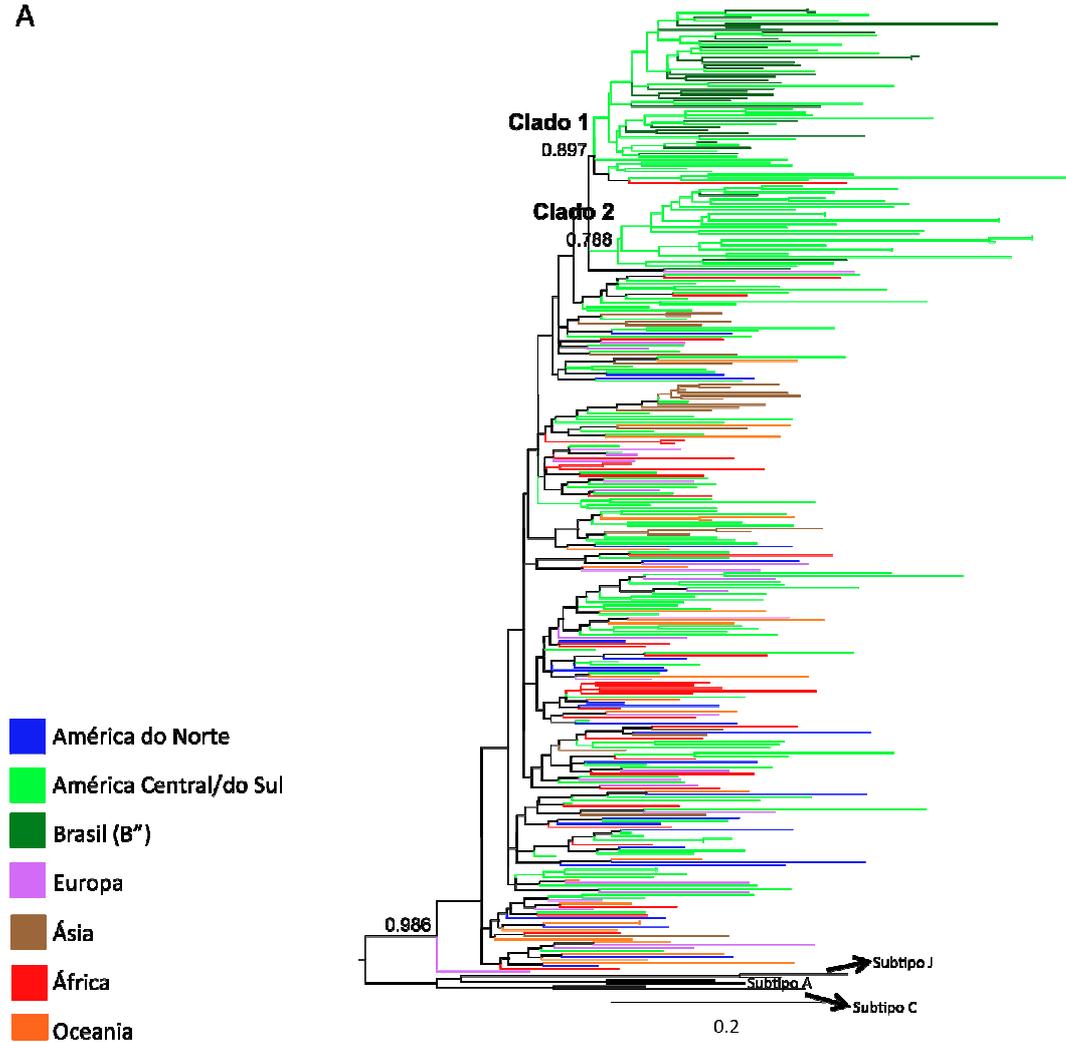
De acordo com a análise filogenética de máxima verossimilhança realizada com seqüências do gene da RT de amostras do subtipo B coletadas no Brasil, assim como aquelas isoladas em outros países da América do Sul, América do Norte, Europa, Ásia, África e Oceania, as linhagens do subtipo B circulantes no Brasil se agruparam com seqüências do subtipo B de várias regiões no mundo, não havendo relação direta e única com algum país ou continente (Figura 23).



Também não foi observada a formação de clados monofiléticos formados exclusivamente por seqüências brasileiras com suporte estatístico significativo. Enfatizando, desta forma, uma panmixia (a mistura completa de seqüências de diferentes áreas geográficas) das linhagens de subtipo B prevalentes nas diferentes regiões geográficas do mundo.

Baseando-se em análises da seqüência de aminoácidos do *loop* V3 de gp120 do envelope, foi caracterizado, no Brasil, um variante do subtipo B, o B'' (Potts, Kalish *et al.*, 1993; Morgado, Sabino *et al.*, 1994). Visando caracterizar a epidemia deste variante no Brasil, seqüências de gp120 foram alinhadas com seqüências do subtipo B circulante em outras regiões do mundo. De acordo com a árvore filogenética gerada por esse alinhamento, pôde-se observar a formação de dois clados monofiléticos brasileiros. O Clado 1 é formado basicamente por variantes B'', com suporte estatístico significativo (0,897) enquanto no Clado 2 (0,788) foram identificadas apenas cinco seqüências com a assinatura GWGR (Figura 24 A). Para verificar a dependência dos clados quanto à presença da assinatura em gp120, uma nova árvore filogenética de máxima verossimilhança foi construída retirando-se os nucleotídeos 934 a 945 de *env*, os quais correspondem à posição da assinatura GWGR. De acordo com essa árvore, apenas o Clado 2 pôde ser identificado. Desse modo caracterizou-se a ocorrência de uma epidemia local do subtipo B do HIV-1 cujas relações filogenéticas independem da presença da assinatura GWGR (Clado 2) e outra epidemia local de variantes B'' (Clado 1) (Figura 24 B). Em nenhuma árvore, os clados monofiléticos 1 e 2 aparecem diretamente relacionados a algum país ou continente. Portanto, não foi possível identificar a origem das linhagens do subtipo B circulantes no Brasil .

A



**Figura 24. Árvore filogenética de máxima verossimilhança do subtipo B do HIV-1 baseada no gene de *gp120* (*env*).** A. Árvore baseada no alinhamento de *gp120* contendo os nucleotídeos 934 a 945 de *env*, que correspondem à posição da assinatura GWGR. B. Árvore baseada no alinhamento de *gp120* excluindo os nucleotídeos 934 a 945 de *env*, que correspondem à posição da assinatura GWGR. Os ramos foram coloridos de acordo com o continente de origem (legenda na esquerda). O número nos pontos de ramificação correspondem ao valor do teste de razão de verossimilhança aproximada tipo SH (aLRT). O comprimento dos ramos em substituições nucleotídicas por sítio segue a escala de acordo com a barra localizada abaixo da árvore. Sequências dos subtipos A (AF004885, AF069670), C (AF67155, AY772699) e J (AF082395, AF082394) foram utilizadas como grupo externo.

B

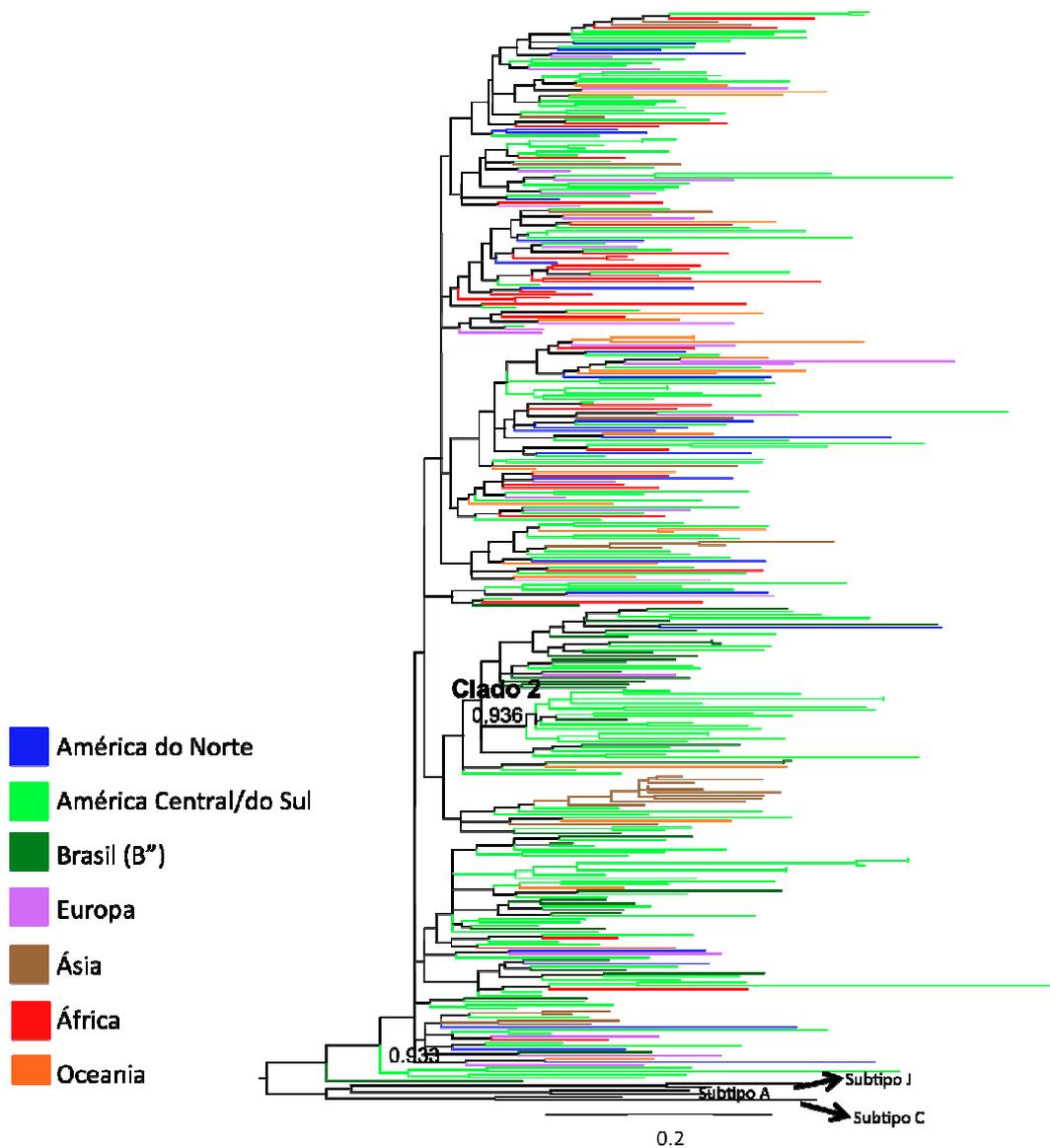


Figura 24. Continuação

## 4.5.2. Subtipo C

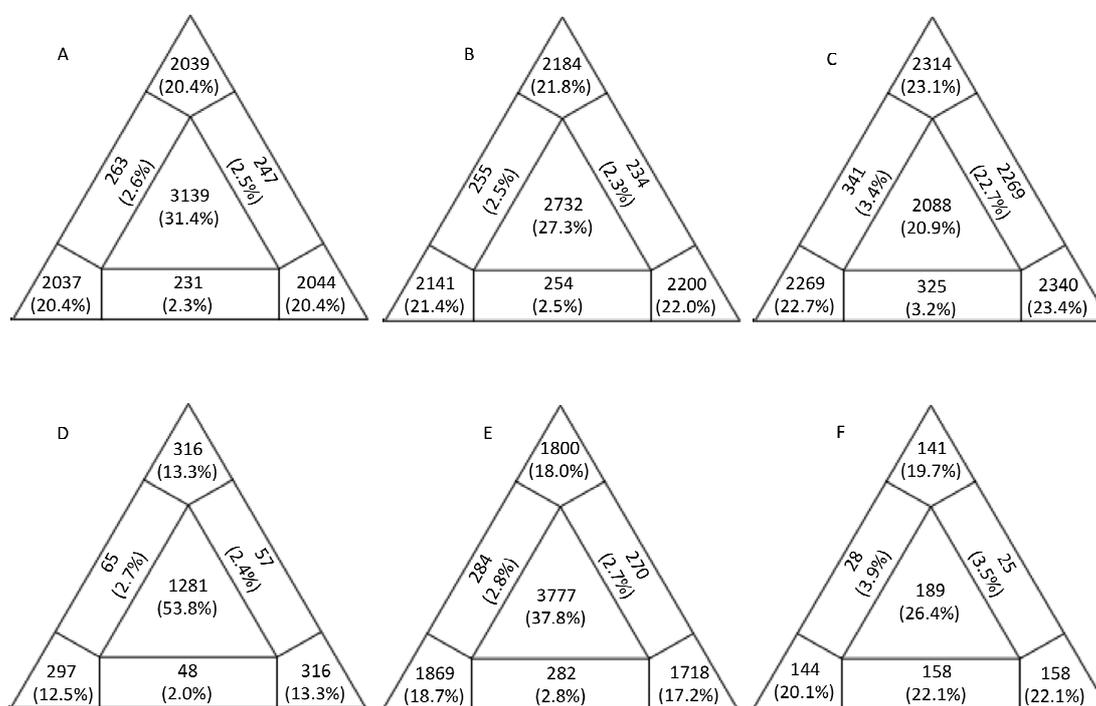
### 4.5.2.1. Mapeamento de Verossimilhança

O mapeamento de máxima verossimilhança de 10.000 quartetos randômicos indicou diferença significativa entre o sinal filogenético dos alinhamentos de p24, RT e gp41 incluindo apenas seqüências brasileiras e daqueles formado por seqüências globais (Figura 25).

A análise incluindo seqüências globais revelaram sinal filogenético do tipo estrela para p24 (31,4%), RT (27,3%) e gp41 (20,9%). A elevada proporção de

quartetos no centro do mapa de verossimilhança é consistente com a dispersão extensiva do subtipo C do HIV-1, o qual é responsável por mais de 50% das infecções por HIV no mundo (Hemelaar, Gouws *et al.*, 2006).

As análises incluindo apenas seqüências brasileiras, por sua vez, indicaram sinal filogenético do tipo estrela ainda maior que aquele descrito para o alinhamento de seqüências globais. A proporção de quartetos no centro do triângulo foi de 53,8% para *p24*, 37,8% para RT e 26,4% para *gp41*. A filogenia do subtipo C do HIV-1 é caracterizada pela emergência simultânea de múltiplas linhagens virais sugerindo uma dispersão exponencial do vírus logo após a sua introdução no país.



**Figura 25. Mapeamento de verossimilhança de linhagens do subtipo C do HIV-1 com distribuição global (triângulos superiores) e restrita ao Brasil (triângulos inferiores).** As análises foram baseadas no gene *p24* (A, 105 seqüências; e D, 20 seqüências), da transcriptase reversa (B, 362 seqüências; e E, 139 seqüências) e *gp41* (*env*) (C, 103 seqüências; e F, 15 seqüências). Os números e porcentagens representam a distribuição detalhada dos quartetos em cada região do mapa. Sete regiões principais podem ser distinguidas: os três vértices, representando sinal filogenético típico de árvore (topologia resolvida); os três lados, representando sinal filogenético do tipo rede (com duas possibilidades de topologia); e o centro, representando sinal filogenético do estrela (topologia não resolvida).

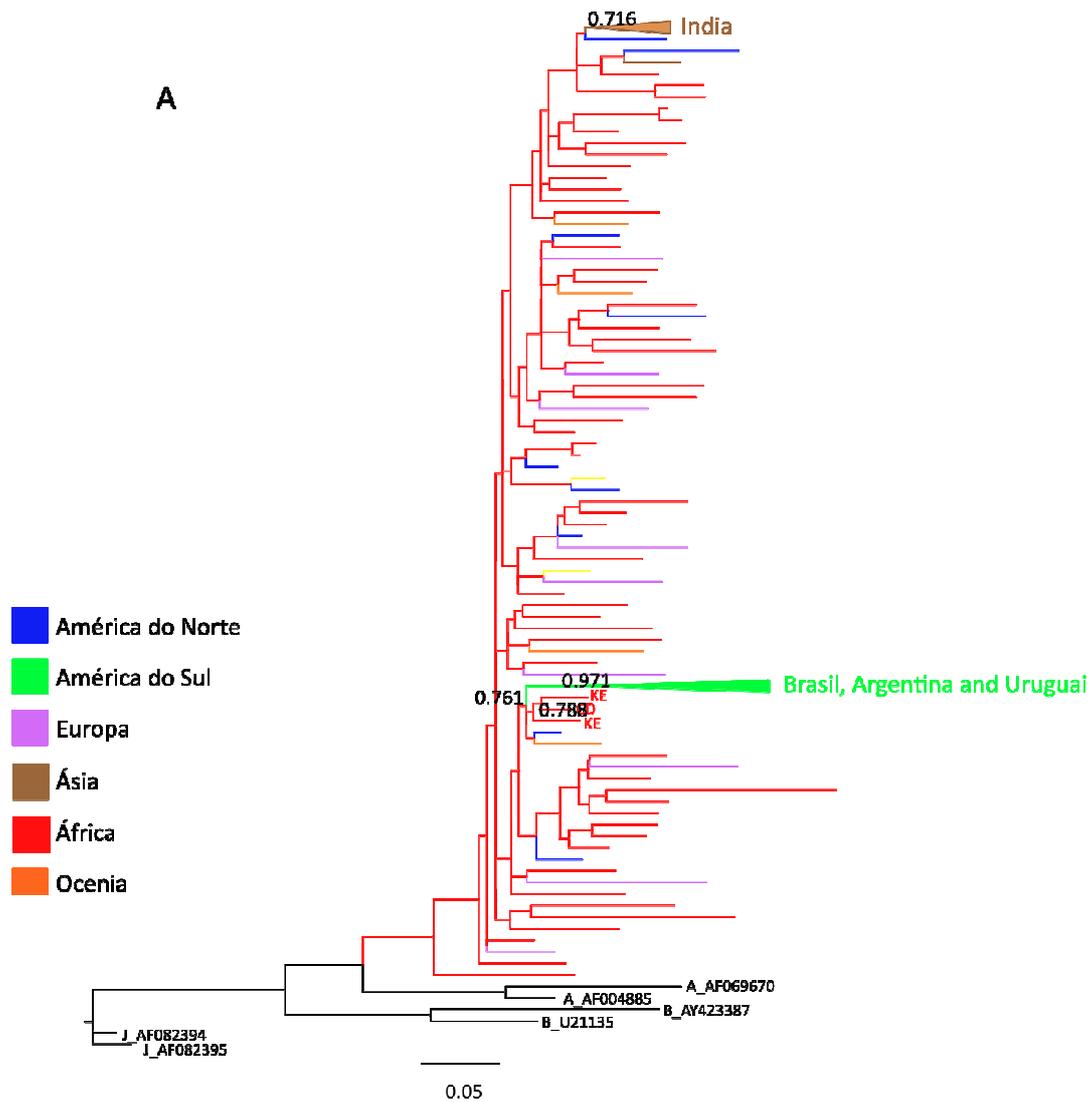
#### 4.5.2.2. Origem do Subtipo C no Brasil

A relação entre as linhagens de subtipo C circulantes no Brasil com aquelas prevalentes em outras regiões geográficas do mundo foi avaliada por meio de análises filogenéticas de máxima verossimilhança baseada nos genes p24, da transcriptase reversa e gp41 (Figura 26).

De acordo com a árvore de p24, seqüências do Brasil, Argentina e Uruguai formaram um clado monofilético (valor de aLRT: 0.971) estreitamente relacionado a seqüências do Kênia e República Democrática do Congo (Figura 26 A). O clado monofilético formado por seqüências da América do Sul (valor de aLRT: 0.789) também foi observado na árvore filogenética de gp41, na qual as amostras sul americanas apresentam relação próxima àquelas de Burundi e República Unida da Tanzânia (Figura 26 C).

Já na árvore baseada no gene da transcriptase reversa, o clado monofilético de amostras da América do Sul (valor de aLRT: 0.908) não aparece relacionado a seqüências de um país específico com suporte estatístico. Duas seqüências brasileiras, DT3SP01 e DO2PR05, agrupam fora do clado monofilético, junto a seqüências de Zâmbia. E, por fim, a seqüência da Venezuela incluída nessa análise também não agrupa junto às seqüências da América do Sul, mas está estreitamente relacionada a uma seqüência da República Democrática do Congo (valor de aLRT: 0.889) (Figura 26 B).

De um modo geral, as análises confirmam a hipótese de que a epidemia do subtipo C na América do Sul foi iniciada pela introdução de seqüências intimamente relacionadas, provavelmente originadas no Leste da África. No entanto, devido aos diferentes padrões de agrupamentos observados nas árvores filogenéticas baseadas nas regiões genômicas de p24, RT e gp41, não foi possível indicar com precisão a origem exata do subtipo C circulante no Brasil.



**Figura 26.** Árvore filogenética de máxima verossimilhança do subtipo C do HIV-1 baseada nos genes *p24* (A), da transcriptase reversa (B) e *gp41 (env)* (C). Os ramos foram coloridos de acordo com o continente de origem (legenda na esquerda de cada árvore). O número nos pontos de ramificação correspondem ao valor do teste de razão de verossimilhança aproximada tipo SH (aLRT). O comprimento dos ramos em substituições nucleotídicas por sítio segue a escala de acordo com a barra localizada abaixo de cada árvore. Sequências dos subtipos A (AF004885, AF069670), C (AF67155, AY772699) e J (AF082395, AF082394) foram utilizadas como grupo externo.

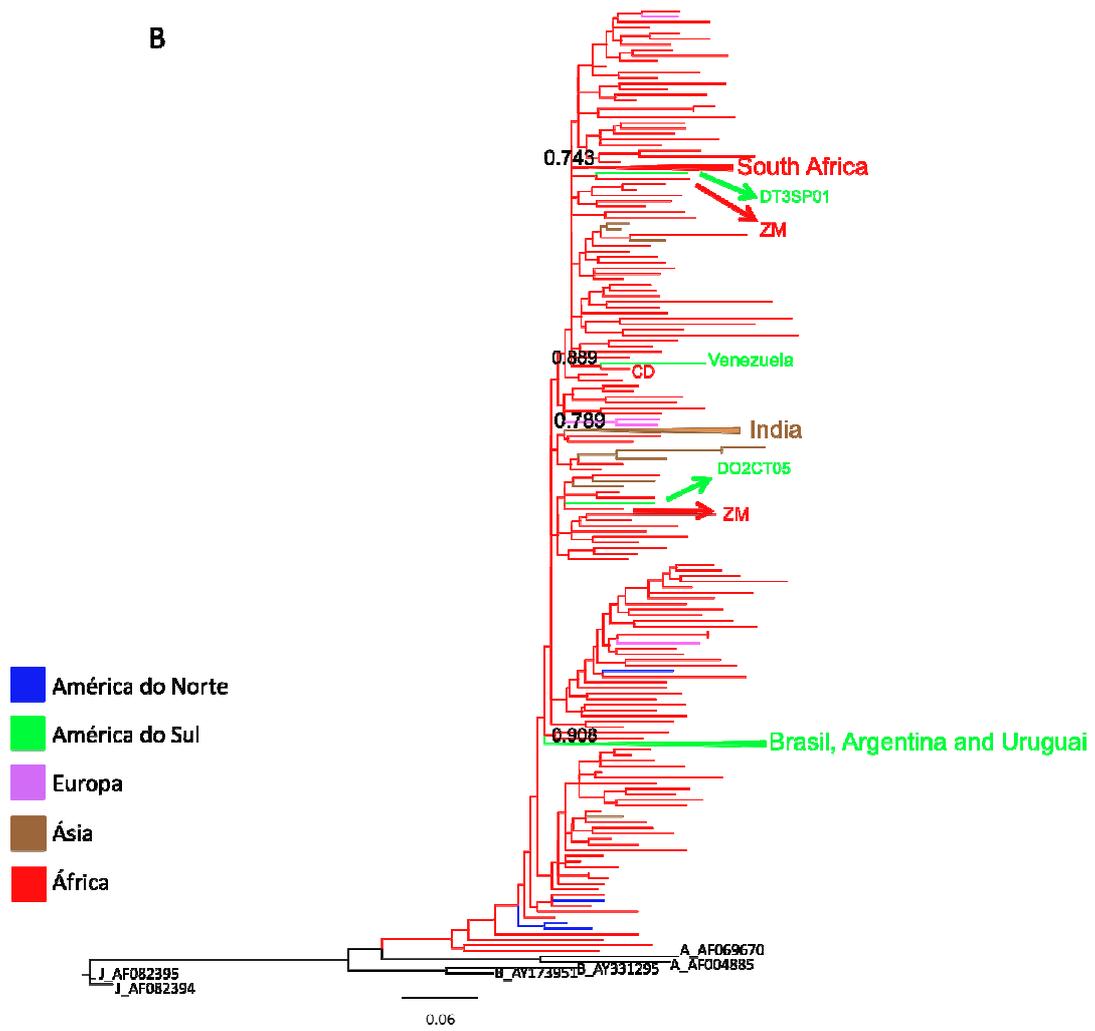
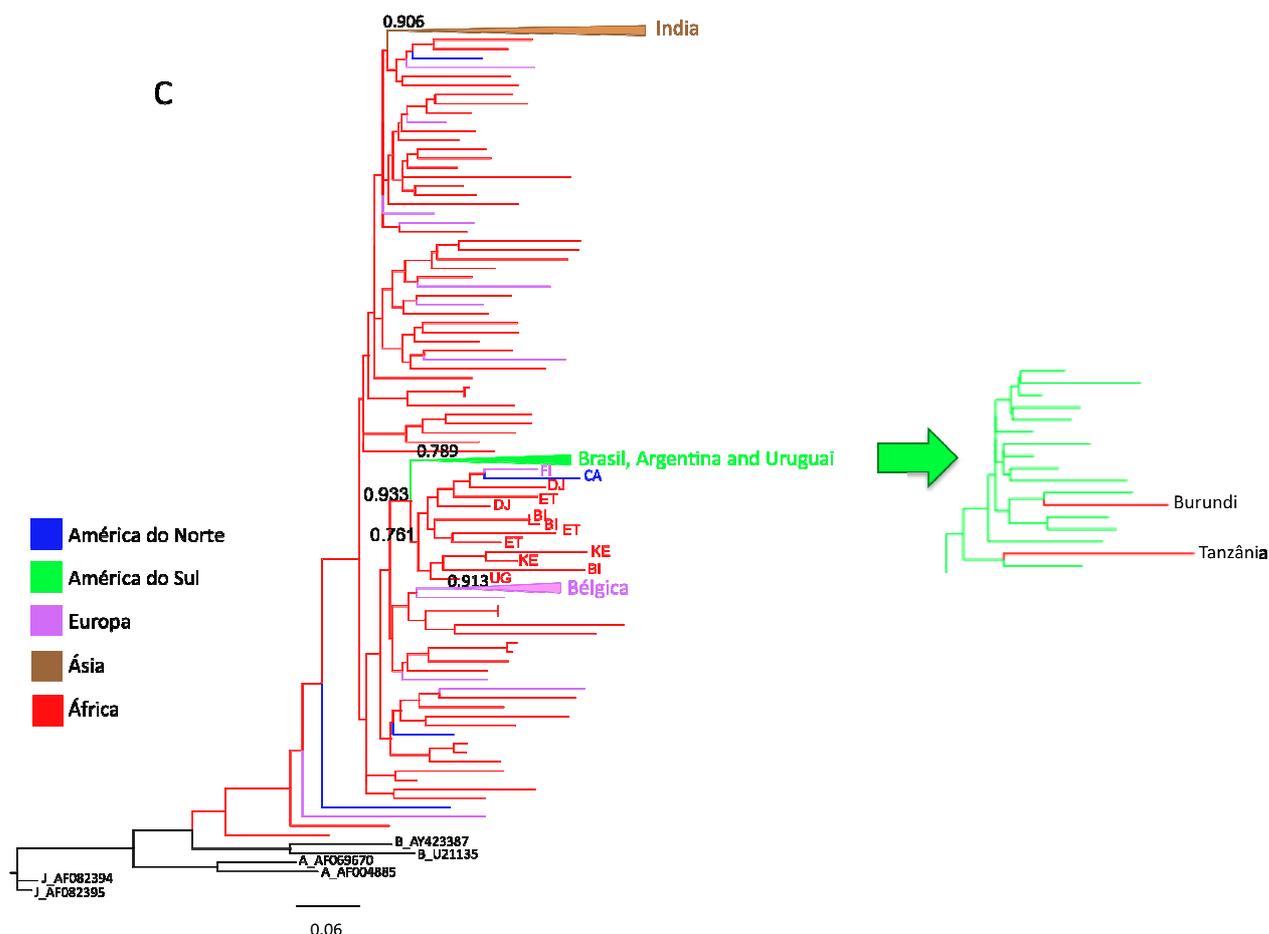


Figura 26. Continuação.



**Figura 26. Continuação**

## 4.6. História Demográfica

### 4.6.1. Subtipo B

A história demográfica do subtipo B do HIV-1 no Brasil foi definida de acordo com análises de coalescência utilizando o alinhamento do gene gp120 das seqüências que formaram os clados 1 e 2 na árvore de máxima verossimilhança, respectivamente. Desse modo, os resultados obtidos refletem apenas a realidade de duas epidemias locais e não podem ser extrapoladas para um contexto em nível nacional. A hipótese nula de relógio molecular estrito foi rejeitada por um Fator de Bayes (FB) de 4,492 para o clado 1 e de 4,546 para o clado 2 (Tabela 5). Aplicando-se relógio molecular relaxado, o modelo populacional que melhor descreveu a dinâmica populacional do subtipo B no Brasil foi o de crescimento exponencial para o clado 1 (Tabela 6) com FB igual a 6,606 quando comparado ao modelo de tamanho populacional constante, e a plotagem Bayesiana para o clado 2 com FB igual a 3,384 quando comparado ao modelo de tamanho populacional constante.

**Tabela 5. Fator de Bayes entre relógio molecular estrito (RE) e relaxado (RR) aplicados ao alinhamento de *gp120 (env)* dos clados 1 e 2 do subtipo B do HIV-1 circulante no Brasil.**

Clado	Relógio Molecular <sup>a</sup>	Verossimilhança Marginal	Fator de Bayes (FB) <sup>b</sup>
1 ( <i>gp120</i> )	RE (H <sub>0</sub> )	-1547,198	
	RR (H <sub>1</sub> )	-1544,952	4,492
	RE (H <sub>0</sub> )	-1547,198	
2 ( <i>gp120</i> )	RR (H <sub>1</sub> )	-1544,952	4,546

<sup>a</sup> O modelo de relógio molecular selecionado está destacado em cinza. H<sub>0</sub>, hipótese nula; H<sub>1</sub>, hipótese alternativa.

<sup>b</sup> 6 > BF > 2 indica evidência positiva contra hipótese nula; 10 > BF > 6 indica forte evidência contra hipótese nula; e BF > 10 indica evidência muito forte contra a hipótese nula

**Tabela 6. Fator de Bayes entre os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana aplicados ao alinhamento de *gp120 (env)* dos clados 1 e 2 do subtipo B do HIV-1 circulante no Brasil.**

Clado	Relógio Molecular	Modelo Demográfico <sup>a</sup>	Verossimilhança Marginal	Fator de Bayes (FB) <sup>b</sup>
1 ( <i>gp120</i> )	RR	Constante	-3243,525	
		Exponencial	-3240,222	6,606
		BSP	-3242,856	1,338
2 ( <i>gp120</i> )	RR	Constante	-1544,952	
		Exponencial	-1545,023	0,142
		BSP	-1546,644	3,384

<sup>a</sup> O modelo demográfico selecionado está destacado em cinza.

<sup>b</sup> 6 > BF > 2 indica evidência positiva contra hipótese nula; 10 > BF > 6 indica forte evidência contra hipótese nula; e BF > 10 indica evidência muito forte contra a hipótese nula

#### 4.6.2. Subtipo C

Análises de coalescência foram realizadas afim de definir o modelo demográfico que melhor se adéqua aos alinhamentos dos genes *p24*, *RT* e *gp41* do subtipo C do HIV-1 prevalente no Brasil. A hipótese nula de relógio molecular estrito foi rejeitada por um FB de 7,968; 10 e 11,996 para *p24*, *RT* e *gp41*, respectivamente (Tabela 7). Aplicando relógio molecular relaxado, o modelo de crescimento populacional exponencial e BSP foi selecionado para *p24* (FB=2,822, para exponencial e FB=3,504, para BSP) e *RT* (FB= 27,342, para exponencial e FB= 31,172, para BSP) quando comparado ao tamanho populacional constante. Apenas para *gp41*, o modelo de tamanho populacional constante representou melhor a dinâmica populacional do subtipo C do HIV-1 no Brasil (FB= 0,142, para exponencial e FB= 0,586, para BSP) (Tabela 8).

**Tabela 7. Fator de Bayes entre relógio molecular estrito (RE) e relaxado (RR) aplicados aos alinhamentos de *p24*, transcriptase reversa (*RT*) e *gp41* (*env*) do subtipo C do HIV-1 circulante no Brasil.**

Gene	Relógio Molecular <sup>a</sup>	Verossimilhança marginal	Fator de Bayes (FB) <sup>b</sup>
<i>p24</i>	RE (H <sub>0</sub> )	-1350,286	7,968
	RR (H <sub>1</sub> )	-1346,302	
RT	RE (H <sub>0</sub> )	-4332,272	10
	RR (H <sub>1</sub> )	-4327,272	
<i>gp41</i>	RR (H <sub>1</sub> )	-1529,565	11,996

<sup>a</sup> O modelo de relógio molecular selecionado está destacado em cinza. H<sub>0</sub>, hipótese nula; H<sub>1</sub>, hipótese alternativa.

<sup>b</sup>  $6 > BF > 2$  indica evidência positiva contra hipótese nula;  $10 > BF > 6$  indica forte evidência contra hipótese nula; e  $BF > 10$  indica evidência muito forte contra a hipótese nula

**Tabela 8. Fator de Bayes entre os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana aplicados aos alinhamentos de *p24*, transcriptase reversa (RT) e *gp41* (*env*) do subtipo C do HIV-1 circulante no Brasil.**

Gene	Relógio		Verossimilhança Marginal	Fator de Bayes (FB) <sup>b</sup>
	Molecular	Modelo Demográfico <sup>a</sup>		
<i>p24</i>	RR	Constante	-1346,302	
		Exponencial	-1344,891	2,822
		BSP	-1344,55	3,504
<i>RT</i>	RR	Constante	-4327,272	
		Exponencial	-4313,601	27,342
		BSP	-4311,686	31,172
<i>gp41</i>	RR	Constante	-1358,949	
		Exponencial	-1358,878	0,142
		BSP	-1358,656	0,586

<sup>a</sup> O modelo de relógio molecular selecionado está destacado em cinza. H<sub>0</sub>, hipótese nula; H<sub>1</sub>, hipótese alternativa.

<sup>b</sup> 6 > BF > 2 indica evidência positiva contra hipótese nula; 10 > BF > 6 indica forte evidência contra hipótese nula; e BF > 10 indica evidência muito forte contra a hipótese nula

## 4.7. Determinação da Taxa Evolutiva e do Tempo do Ancestral Comum Mais Recente

### 4.7.1. Subtipo B

Para estimar a taxa evolutiva ( $\mu$ , substituições de nucleotídeos por sítio por ano) e o tempo do ancestral comum mais recente (Tmrca, anos) do subtipo B do HIV-1 no Brasil, foram utilizados os alinhamentos contendo as seqüências do subtipo B que formaram os cladogramas monofiléticos 1 e 2 na análise de máxima verossimilhança e seqüências de referência dos subtipos do HIV-1. Uma vez que o Tmrca do grupo M do HIV-1 é bem estabelecido como tendo ocorrido na década de 30 (Korber, Muldoon *et al.*, 2000; Salemi, Strimmer *et al.*, 2001), e a origem do subtipo B na década de 60 (Deng, Liu *et al.*, 2008), os subtipos do HIV-1 foram utilizados como controle para verificar a veracidade dos resultados obtidos.

O modelo de crescimento populacional constante, aplicando taxa evolutiva normal como premissa, apresentou os resultados mais fidedignos, de acordo com o obtido

para o grupo M e subtipo B do HIV-1 para ambos os clados. De acordo com essa análise, a taxa evolutiva do subtipo B do HIV-1 circulante no Brasil foi  $5,9 \times 10^{-3}$  (95% HPD =  $1,4 \times 10^{-3} - 1,5 \times 10^{-2}$ ) substituições sítios<sup>-1</sup> ano<sup>-1</sup>, para o clado 1; e  $5,7 \times 10^{-3}$  (95% HPD =  $1,2 \times 10^{-3} - 1,6 \times 10^{-2}$ ) substituições sítios<sup>-1</sup> ano<sup>-1</sup>, para o clado 2. A entrada da linhagem do clado 1 no Brasil foi estimada em 1972 (95% HPD = 1957-1982) e a do clado 2, em 1973 (95% HPD = 1951-1984). De acordo com os cálculos baseados no clado 1, a origem do subtipo B global ocorreu em 1962 (95% HPD = 1941-1977) e a do grupo M, em 1928 (95% HPD = 1882-1958). Os valores obtidos para o clado 2 foram 1964 (95% HPD = 1938-1980) para origem do subtipo B global e 1930 (95% HPD = 1871-1961) para a origem do grupo M do HIV-1 (Tabela 9).

**Tabela 9. Tempo do ancestral comum mais recente (Tmrca) e taxa evolutiva do subtipo B aplicando relógio molecular relaxado e os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana.**

Clado	Relógio Molecular	Modelo de crescimento populacional	Tmrca do subtipo B no Brasil	Tmrca do subtipo B global	Tmrca do grupo M	Taxa evolutiva	Tmrca do subtipo B no Brasil (taxa) <sup>a</sup>	Tmrca do subtipo B global (taxa) <sup>a</sup>	Tmrca do grupo M (taxa) <sup>a</sup>	Taxa evolutiva (taxa) <sup>a</sup>
Clado 1 (gp120)	RR	Constante	1973 (1955-1983)	1964 (1938-1978)	1931 (1879-1963)	$6,2 \times 10^{-3}$ ( $1,3 \times 10^{-3} - 1,6 \times 10^{-2}$ )	1972 (1957-1982)	1962 (1941-1977)	1928 (1882-1958)	$5,9 \times 10^{-3}$ ( $1,4 \times 10^{-3} - 1,4 \times 10^{-2}$ )
		Exponencial	1924 (1732-1977)	1909 (1686-1974)	1874 (1565-1964)	$2,3 \times 10^{-3}$ ( $2,0 \times 10^{-4} - 6,9 \times 10^{-3}$ )	1938 (1843-1971)	1931 (1826-1969)	1908 (1765-1959)	$2,8 \times 10^{-3}$ ( $3,0 \times 10^{-4} - 6,8 \times 10^{-3}$ )
		BSP	1932 (1756-1973)	1921 (1715-1972)	1861 (1494-1952)	$2,3 \times 10^{-3}$ ( $1,0 \times 10^{-4} - 5,8 \times 10^{-3}$ )	1948 (1875-1973)	1939 (1855-1969)	1893 (1745-1953)	$3,1 \times 10^{-3}$ ( $3,0 \times 10^{-4} - 6,8 \times 10^{-3}$ )
Clado 2 (gp120)	RR	Constante	1974 (1942-1987)	1966 (1924-1982)	1934 (1846-1969)	$5,8 \times 10^{-3}$ ( $2,0 \times 10^{-4} - 1,8 \times 10^{-2}$ )	1973 (1951-1984)	1964 (1938-1980)	1930 (1871-1961)	$5,7 \times 10^{-3}$ ( $1,2 \times 10^{-3} - 1,6 \times 10^{-2}$ )
		Exponencial	1924 (1732-1977)	1909 (1686-1974)	1874 (1565-1964)	$2,3 \times 10^{-3}$ ( $2,0 \times 10^{-4} - 6,9 \times 10^{-3}$ )	1942 (1711-1961)	1931 (1798-1974)	1904 (1828-1978)	$3,0 \times 10^{-3}$ ( $3,0 \times 10^{-4} - 8,0 \times 10^{-3}$ )
		BSP	1939 (1594-1979)	1926 (1503-1975)	1878 (1179-1964)	$2,3 \times 10^{-3}$ ( $0 - 7,0 \times 10^{-3}$ )	1957 (1885-1979)	1948 (1861-1977)	1916 (1772-1966)	$3,6 \times 10^{-3}$ ( $4,0 \times 10^{-4} - 8,6 \times 10^{-3}$ )

<sup>a</sup> Taxa evolutiva =  $5 \pm 2 \times 10^{-3}$  substituições nucleotídicas/sítio/ano (Bello, Eyer-Silva *et al.*, 2007), utilizada como premissa.

#### 4.7.2. Subtipo C

Para os cálculos de  $\mu$  e Tmrcas das linhagens do subtipo C circulante no Brasil, também foram acrescentadas às seqüências do subtipo C brasileiras, seqüências de referência dos subtipos do HIV, considerando a origem do Grupo M do HIV-1 na década de 30 (Korber, Muldoon *et al.*, 2000; Salemi, Strimmer *et al.*, 2001) e o subtipo C global em meados da década de 60 (Travers, Clewley *et al.*, 2004). O modelo de crescimento populacional constante, estabelecendo como premissa a taxa evolutiva com distribuição normal, de acordo com o especificado na metodologia, apresentou os resultados mais reais conforme o obtido para o grupo M e subtipo C global do HIV-1.

A taxa evolutiva do subtipo C do HIV-1 prevalente no Brasil foi  $1,9 \times 10^{-3}$  (95% HPD=  $7,0 \times 10^{-4} - 3,6 \times 10^{-3}$ ) substituições sítios<sup>-1</sup> ano<sup>-1</sup>,  $2,1 \times 10^{-3}$  (95% HPD=  $7,0 \times 10^{-4} - 3,6 \times 10^{-3}$ ) substituições sítios<sup>-1</sup> ano<sup>-1</sup>, e  $1,5 \times 10^{-3}$  (95% HPD=  $7,0 \times 10^{-4} - 2,4 \times 10^{-3}$ ) substituições sítios<sup>-1</sup> ano<sup>-1</sup>, para os genes p24, transcriptase reversa e gp41, respectivamente.

Baseando-se nas análises de gp41, o Tmrcas do subtipo C do HIV-1 no Brasil foi estimado em 1976 (95% HPD= 1961-1984). De acordo com essas análises, o Tmrcas do grupo M do HIV-1 teria sido em 1933 (95% HPD= 1894-1958) e, do subtipo C global, em 1965 (95% HPD= 1943-1977). Essas datas estão de acordo com o previamente publicado por Korber *et al* (2000), Salemi (2001) e Travers *et al* (2004) (Tabela 10).

As análises baseadas em p24 e transcriptase reversa dataram a origem do subtipo C na América do Sul em 1986 (95% HPD= 1978-1990) e 1984 (95% HPD= 1978-1989), respectivamente. No entanto, as datas da origem do grupo M do HIV-1 e do subtipo C global foram aproximadamente no início da década de 50 e final de da década de 70, respectivamente, não correspondendo àquelas publicadas pela literature (Tabela 10). Dessa forma, as datas estimadas por gp41 foram consideradas mais fidedignas.

**Tabela 10. Tempo do ancestral comum mais recente (Tmrca) e taxa evolutiva do subtipo C aplicando relógio molecular relaxado e os modelos demográficos de tamanho populacional constante, crescimento populacional exponencial e plotagem Bayesiana.**

Gene	Relógio Molecular	Modelo de crescimento populacional	Tmrca do subtipo C no Brasil	Tmrca do subtipo C global	Tmrca do grupo M	Taxa evolutiva	Tmrca do subtipo C no Brasil (taxa)	Tmrca do subtipo C global (taxa)	Tmrca do grupo M (taxa)	Taxa evolutiva (taxa)
p24	RR	Constante	1988 (1983-1991)	1975 (1962-1984)	1947 (1917-1966)	$2,3 \times 10^{-3}$ ( $8,0 \times 10^{-4}$ - $4,8 \times 10^{-3}$ )	1987 (1979-1991)	1972 (1956-1981)	1938 (1904-1959)	$1,9 \times 10^{-3}$ ( $7,0 \times 10^{-4}$ - $3,6 \times 10^{-3}$ )
		Exponencial	1961 (1839-1991)	1935 (1740-1981)	1875 (1519-1959)	$8,0 \times 10^{-4}$ ( $1,0 \times 10^{-4}$ - $1,0 \times 10^{-3}$ )	1972 (1936-1988)	1951 (1898-1977)	1906 (1802-1951)	$1,0 \times 10^{-3}$ ( $3,0 \times 10^{-4}$ - $2,1 \times 10^{-3}$ )
		BSP	1848 (0-1991)	1763 (0-1980)	1511 (0-1967)	$2,0 \times 10^{-4}$ ( $0-9,0 \times 10^{-4}$ )	1905 (1153-1979)	1905 (1153-1979)	1805 (285-1959)	$5,0 \times 10^{-4}$ ( $0-1,5 \times 10^{-3}$ )
RT	RR	Constante	1985 (1980-1989)	1979 (1971-1986)	1954 (1938-1967)	$2,0 \times 10^{-3}$ ( $7,0 \times 10^{-4}$ - $4,0 \times 10^{-3}$ )	1985 (1979-1989)	1978 (1970-1985)	1952 (1936-1965)	$1,9 \times 10^{-3}$ ( $7,0 \times 10^{-4}$ - $3,6 \times 10^{-3}$ )
		Exponencial	1873 (1379-1988)	1853 (1242-1983)	1794 (966-1967)	$3,0 \times 10^{-4}$ ( $0-9,0 \times 10^{-4}$ )	1934 (1746-1987)	1923 (1706-1981)	1888 (1593-1967)	$5,0 \times 10^{-4}$ ( $0-1,5 \times 10^{-4}$ )
		BSP	1717 (0-1974)	1654 (0-1967)	1335 (0-1927)	$1,0 \times 10^{-4}$ ( $0-4,0 \times 10^{-4}$ )	1880 (1541-1977)	1852 (1443-1970)	1711 (945-1931)	$2,0 \times 10^{-4}$ ( $0-7,0 \times 10^{-4}$ )
gp41	RR	Constante	1974 (1955-1984)	1962 (1934-1977)	1928 (1876-1956)	$1,4 \times 10^{-3}$ ( $6,0 \times 10^{-4}$ - $2,3 \times 10^{-3}$ )	1976 (1961-1984)	1965 (1944-1977)	1933 (1895-1958)	$1,5 \times 10^{-3}$ ( $7,0 \times 10^{-4}$ - $2,4 \times 10^{-3}$ )
		Exponencial	1965 (1920-1982)	1950 (1888-1975)	1950 (1888-1975)	$1,1 \times 10^{-3}$ ( $3,0 \times 10^{-4}$ - $1,8 \times 10^{-3}$ )	1972 (1952-1984)	1961 (1934-1977)	1932 (1882-1958)	$1,4 \times 10^{-3}$ ( $6,0 \times 10^{-4}$ - $2,2 \times 10^{-3}$ )
		BSP	1950 (1761-1983)	1931 (1671-1975)	1875 (1412-1960)	$7,0 \times 10^{-4}$ ( $3,0 \times 10^{-4}$ - $1,8 \times 10^{-3}$ )	1967 (1931-1983)	1955 (1905-1976)	1919 (1828-1957)	$1,1 \times 10^{-3}$ ( $4,0 \times 10^{-4}$ - $2,0 \times 10^{-3}$ )

<sup>a</sup> Taxa evolutiva =  $13 \pm 5 \times 10^{-4}$  substituições nucleotídicas/sítio/ano, para p24 (Lemey, Pybus *et al.*, 2004),  $16 \pm 6 \times 10^{-4}$  substituições nucleotídicas/sítio/ano, para RT (Bello, Passaes *et al.*, 2008), e  $31 \pm 9 \times 10^{-4}$  substituições nucleotídicas/sítio/ano, para gp41 (Lemey, Pybus *et al.*, 2004), utilizadas como premissa.

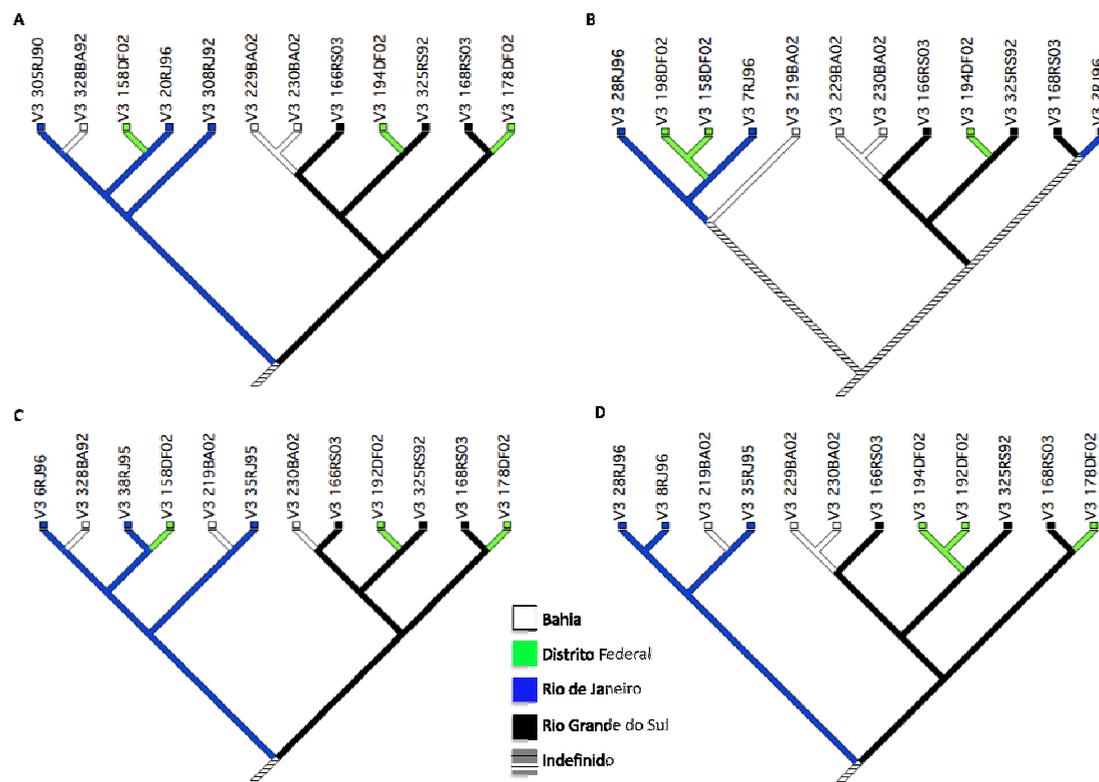
## 4.8. Análises Filogeográficas

### 4.8.1. Subtipo B

A análise do fluxo gênico do subtipo B do HIV-1 nos diferentes estados brasileiros foi baseada na região genômica de *gp120* (*env*), incluindo apenas as amostras que formaram o clado monofilético 1. Foram analisados quatro alinhamentos contendo, cada um, três seqüências selecionadas aleatoriamente da Bahia, Distrito Federal, Rio de Janeiro e Rio Grande do Sul. Não foi possível realizar análises filogeográficas para o clado 2 por limitações de amostragens.

Nas quatro árvores, a RMP dos caracteres ancestrais indicou o Rio de Janeiro e o Rio Grande do Sul como os dois lugares a partir dos quais as linhagens do subtipo B circulantes em outros estados brasileiros se originaram, sendo considerados, portanto, dois pontos de entrada do subtipo B no Brasil (Figura 27).

De acordo com o teste de metapopulação, a hipótese nula de panmixia não foi rejeitada ( $p > 0.01$ ), em todos os alinhamentos analisados, indicando a mistura de linhagens virais prevalentes em diferentes regiões geográficas (Tabela 11).



**Figura 27. Padrões filogeográficos do subtipo B do HIV-1 no Brasil.** Quatro análises filogeográficas representativas utilizando genealogia enraizada baseada no gene *gp120* (*env*) de 12 seqüências da Bahia, Distrito Federal, Rio de Janeiro e Rio Grande do Sul (três seqüências de cada estado) que compunham o clado monofilético 1 presente na análise filogenética de máxima verossimilhança. A reconstrução mais parcimônia (RMP) do estado de origem de cada nó interno (seqüência ancestral) da árvore está indicado pela cor do ramo correspondente de acordo com a legenda. Ramos indefinidos indicam múltiplas RMP.

**Tabela 11. Análise de metapopulação do subtipo B do HIV-1 circulante no Brasil.**

Alinhamento <sup>a</sup>	Fluxo gênico observado <sup>b</sup>	95% I.C. <sup>c</sup>	valor de $p$ <sup>d</sup>
1	6	6-9	0,1802
2	6	6-8	0,1726
3	7	6-9	0,5885
4	5	6-9	0,0251

<sup>a</sup> Cada alinhamento incluiu 12 seqüências da Bahia, Distrito Federal, Rio de Janeiro e Rio Grande do Sul (três seqüências de cada estado) que compunham o clado monofilético 1 presente na análise filogenética de máxima verossimilhança.

<sup>b</sup> Número de migrações observadas na árvore Bayesiana.

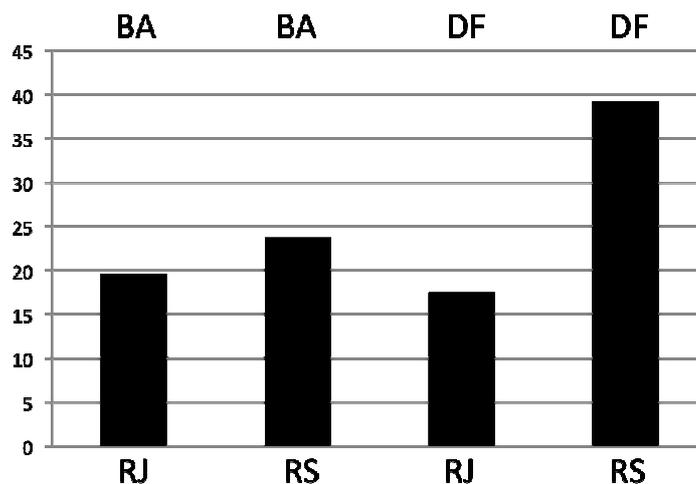
<sup>c</sup> Intervalo de confiança de 95% da distribuição de migrações de 10.000 árvores randômicas.

<sup>d</sup>  $p > 0,01$  indica subdivisão populacional não significativa entre diferentes estados.

Por fim, é notável a presença de um fluxo gênico assimétrico do subtipo B, com infecções viras dispersando especialmente do Rio de Janeiro e Rio Grande do Sul para o Centro-Oeste e Nordeste do Brasil (Figura 28), de modo que todas as migrações observadas apresentaram sentido Sul-Norte (Figura 29)



**Figura 28. Padrões de migração do subtipo B do HIV-1 no Brasil.** O fluxo gênico do subtipo B no Brasil foi avaliado por meio de quatro análises filogeográficas representativas baseadas no gene *gp120 (env)* de 12 seqüências da Bahia, Distrito Federal, Rio de Janeiro e Rio Grande do Sul (três seqüências de cada estado) que compunham o clado monofilético 1 presente na análise filogenética de máxima verossimilhança. As setas representam a direção do fluxo gênico do subtipo B. A largura das setas é proporcional à média do número relativo de eventos de migração entre os diferentes estados brasileiros observados nas genealogias Bayesianas da Figura 26.

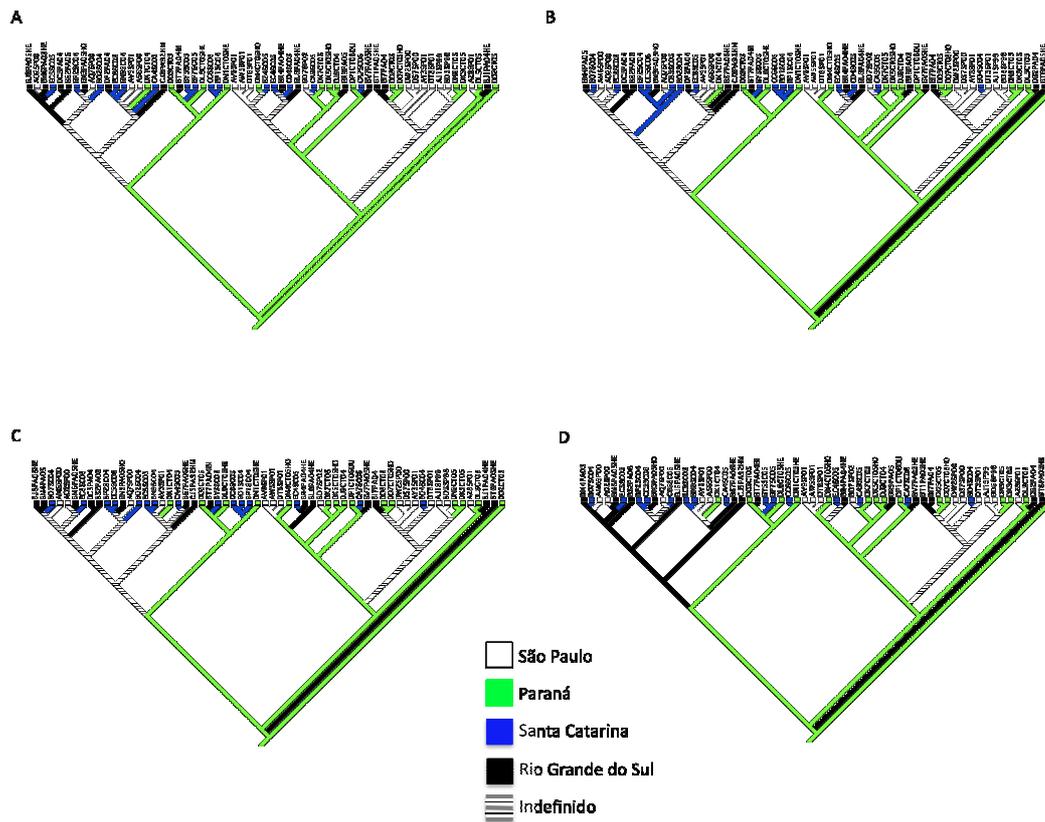


**Figura 29. Rede de fluxo gênico do subtipo B do HIV-1 entre os diferentes estados brasileiros.** As barras representam a rede de fluxo gênico viral entre o par de estados ao longo do eixo Norte-Sul [(média de fluxo do estado A para o estado B) - (média de fluxo do estado B para o estado A)]. BA, Bahia; DF, Distrito Federal; RJ, Rio de Janeiro e RS, Rio Grande do Sul.

#### 4.8.2. Subtipo C

A análise da migração do subtipo C do HIV-1 entre os diferentes estados do Brasil foi baseada no gene da RT por ser o único alinhamento com seqüências suficientes de diferentes estados brasileiros. Quatro alinhamentos, contendo cada um 15 seqüências selecionadas randomicamente de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul, foram analisados. A RMP dos caracteres ancestrais nas quatro árvores mostrou, consistentemente, que o Paraná foi o estado de entrada do subtipo C no Brasil (Figura 30).

De acordo com o teste de metapopulação, a hipótese nula de panmixia não pôde ser rejeitada ( $p > 0.01$ ), em todos os alinhamentos analisados, indicando a mistura de linhagens virais de diferentes regiões geográficas (Tabela 12).



**Figura 30. Padrões filogeográficos do subtipo C do HIV-1 no Brasil.** Quatro análises filogeográficas representativas utilizando genealogia enraizada baseada no gene da transcriptase reversa de 60 seqüências, escolhidas randomicamente, de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (15 seqüências de cada estado). A reconstrução mais parcimônia (RMP) do estado de origem de cada nó interno (seqüência ancestral) da árvore está indicado pela cor do ramo correspondente de acordo com a legenda. Ramos indefinidos indicam múltiplas RMP.

**Tabela 12. Análise de metapopulação do subtipo C do HIV-1 circulante no Brasil.**

Alinhamento <sup>1</sup>	Fluxo gênico observado <sup>2</sup>	95% I.C. <sup>3</sup>	valor de $p$ <sup>4</sup>
1	30	30-37	0,07
2	30	30-37	0,07
3	29	30-37	0,03
4	31	30-37	0,15

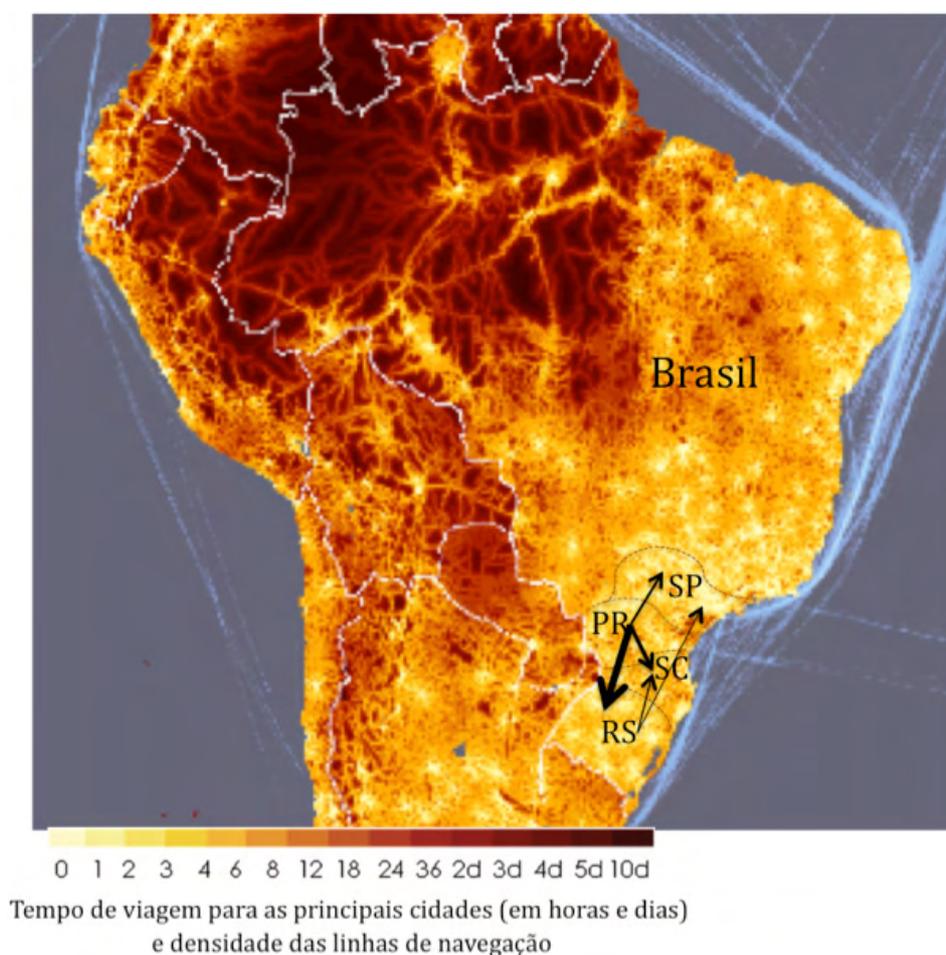
<sup>a</sup> Cada alinhamento incluiu 60 seqüências de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (15 seqüências de cada estado).

<sup>b</sup> Número de migrações observadas na árvore Bayesiana.

<sup>c</sup> Intervalo de confiança de 95% da distribuição de migrações de 10.000 árvores randômicas.

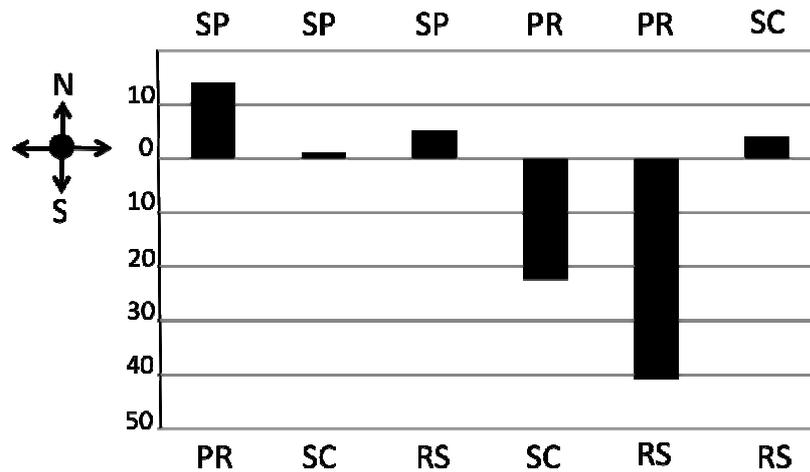
<sup>d</sup>  $p > 0,01$  indica subdivisão populacional não significativa entre diferentes estados.

O mapa de acessibilidade (Figura 31) mostra o tempo estimado de deslocamento dos principais centros urbanos (> 500.000 habitantes) para qualquer lugar de interesse no mapa. Baseando-se nisso, a interconexão nas cidades do Sul e Sudeste é bastante clara, ao contrário das regiões Norte, Nordeste e Centro-Oeste. Esse pode ser um dos aspectos relacionados ao fato de que o subtipo C do HIV-1 está basicamente restrito ao Sul do país. Além disso, é evidente a existência de uma migração altamente assimétrica do subtipo C entre os diferentes estados brasileiros, como representado pelas setas na Figura 30. O Paraná aparece como o ponto central a partir do qual as linhagens virais se espalham radialmente para São Paulo, Santa Catarina e Rio Grande do Sul. Um menor número de migrações do Rio Grande do Sul para Santa Catarina e São Paulo também foi identificada.



**Figura 31. Mapa de acessibilidade e padrões de migração do subtipo C do HIV-1 no Brasil.** O mapa, adaptado de (Nelson, 2008), ilustra o tempo de viagem esperado, de acordo com as cores da legenda, dos grandes centros urbanos para um local de interesse por meio de locomoção terrestre ou aquática. As setas representam a direção do fluxo gênico do subtipo C no Brasil. A largura das setas é proporcional à média do número relativo de eventos de migração entre os diferentes estados brasileiros observados nas genealogias Bayesianas da Figura 29.

Ao contrário do observado para o subtipo B, houve um maior fluxo gênico do subtipo C com direção Norte-Sul (do Paraná para Santa Catarina e Rio Grande do Sul), o qual representou 63% das migrações observadas, que no sentido Sul-Norte (do Rio Grande do Sul para Santa Catarina e São Paulo e do Paraná para São Paulo), que correspondeu a 37% do fluxo gênico (Figura 32).



**Figura 32. Rede de fluxo gênico do subtipo C do HIV-1 entre os diferentes estados brasileiros.** As barras representam a rede de fluxo gênico viral entre o par de estados ao longo do eixo Norte-Sul [(média de fluxo do estado A para o estado B) - (média de fluxo do estado B para o estado A)]. PR, Paraná; RS, Rio Grande do Sul; SC, Santa Catarina; e SP, São Paulo.