



UNIVERSIDADE FEDERAL DE SANTA CATARINA

CENTRO DE CIÊNCIAS BIOLÓGICAS

DEPARTAMENTO DE MICROBIOLOGIA, IMUNOLOGIA E PARASITOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E BIOCÊNCIA

Hegger Machado Fritsch

**DINÂMICA DE TRANSMISSÃO E CARACTERIZAÇÃO MOLECULAR DA
EPIDEMIA DE HIV-1 EM PACIENTES EM FALHA TERAPÊUTICA DE SANTA
CATARINA, BRASIL: UMA ABORDAGEM FILOGENÉTICA**

Florianópolis

2020

Hegger Machado Fritsch

**DINÂMICA DE TRANSMISSÃO E CARACTERIZAÇÃO MOLECULAR DA
EPIDEMIA DE HIV-1 EM PACIENTES EM FALHA TERAPÊUTICA DE SANTA
CATARINA, BRASIL: UMA ABORDAGEM FILOGENÉTICA**

Dissertação submetida ao Programa de Pós Graduação em Biotecnologia e Biociência da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestra em Biotecnologia e Biociências. Área de concentração: Microbiologia e Parasitologia.

Orientador: Prof. Dr. Aguinaldo Roberto Pinto

Co-orientador: Prof. Dr. Tiago Gräf

Florianópolis

2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Fritsch, Hegger Machado

Dinâmica de transmissão e caracterização molecular da
epidemia de HIV-1 em pacientes em falha terapêutica de
Santa Catarina, Brasil: uma abordagem filogenética. /
Hegger Machado Fritsch ; orientador, Aguinaldo Roberto
Pinto, coorientador, Tiago Gräf, 2020.

128 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro de Ciências Biológicas, Programa de Pós
Graduação em Biotecnologia e Biociências, Florianópolis,
2020.

Inclui referências.

1. Biotecnologia e Biociências. 2. HIV. 3.
Epidemiologia Molecular. 4. Redes de Transmissão. 5.
Dispersão Viral. I. Pinto, Aguinaldo Roberto. II. Gräf,
Tiago. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Biotecnologia e Biociências.
IV. Título.

Hegger Machado Frtisch

**DINÂMICA DE TRANSMISSÃO E CARACTERIZAÇÃO MOLECULAR DA
EPIDEMIA DE HIV-1 EM PACIENTES EM FALHA TERAPÊUTICA DE SANTA
CATARINA, BRASIL: UMA ABORDAGEM FILOGENÉTICA.**

O presente trabalho em nível de Mestrado foi avaliado e aprovado por banca
examinadora composta pelos seguintes membros:

Dr. Guilherme de Toledo e Silva
Universidade Federal de Santa Catarina

Prof. Dr. Glauber Wagner
Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que
foi julgado adequado para obtenção do título de Mestra em Biotecnologia e
Biociências.

Prof. Dr. Glauber Wagner
Coordenador do Programa de Biotecnologia e Biociências

Prof. Dr. Aguinaldo Roberto Pinto
Orientador

Este trabalho foi realizado no Laboratório de Imunologia Aplicada/CCB/UFSC, em colaboração com o Laboratório de Virologia Molecular/Instituto de Biologia/UFRJ, Instituto Gonçalo Moniz – Fiocruz/BA e parceria com o Departamento de ISTs/AIDS/Ministério da Saúde, sob orientação dos Prof. Dr. Aguinaldo Roberto Pinto e Prof. Dr. Tiago Gräf, com recursos financeiros provenientes da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

*“Lembre-se de olhar para o alto, para as estrelas,
e não para baixo, para os seus pés.”*

Stephen Hawking

*“Cada pessoa deve trabalhar para o seu aperfeiçoamento e,
ao mesmo tempo, participar da responsabilidade
coletiva para toda a humanidade.”*

Marie Curie

“A persistência é o caminho do êxito.”

Charlie Chaplin

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e força para chegar ao final desta caminhada.

A minha amada avó Zilah Machado que não pode estar presente em mais essa conquista, mas cuja força me motiva e impulsiona desde sempre. Saudades.

Agradeço aos meus pais Ana Lúcia e Paulo Roberto pelo apoio incondicional em todos os momentos da minha vida. Por serem meu exemplo de força, dedicação, amor, respeito e, principalmente, por batalharem e sonharem meus sonhos comigo. Desculpem todos os momentos de ausência e abdições que tivemos que ter para que esse trabalho fosse possível. Obrigada por todo o incentivo e apoio nessa caminhada. Este trabalho é para vocês (e por vocês)!!

A dona Hilda Ribeiro, por todo amor e dedicação em cuidar de mim. Obrigada por sempre torcer e comemorar minhas vitórias.

Agradeço a minha família e amigos que, nesses anos, me ensinaram que para estar junto não é preciso estar perto fisicamente. Obrigada por incentivarem meus sonhos, mesmo que isso signifique uma distância maior e um convívio menor. Em cada momento dessa caminhada levei vocês no meu coração. Em especial as amigas Franciele Giongo, Thainá Garbino e Nathália Kersting por todas as conversas, incentivo e risadas ao longo da vida, sem dúvidas foram os maiores presentes que a biomedicina e o Rosário me trouxeram. Impossível deixar de agradecer e expressar meu amor pelos meus dois pequenos: Isabelly e Otávio. Obrigada por entenderem a minha ausência no crescimento de vocês. A saudade é imensa.

Ao meu orientador Dr. Aguinaldo Roberto Pinto, por ter aceitado mais uma orientada em épocas de recursos escassos, abraçado o desafio em desenvolver um projeto de mestrado totalmente *in silico* e ter me incentivado a aprimorar meus conhecimentos em outras instituições. Certamente esse apoio foi fundamental para que o trabalho pudesse ser concluído e para meu crescimento pessoal e profissional. Só tenho gratidão pela oportunidade e pela confiança. Muito obrigada!

Ao meu orientado Dr. Tiago Gräf, não tenho palavras para descrever esses anos de parceria. Obrigada por sempre confiar em mim e no meu trabalho, por aceitar me orientar a distância, mais uma vez, e ser sempre tão presente para me auxiliar nas análises, nos inúmeros problemas que sempre surgem, em esclarecer as dúvidas e fazer eu criar mais inúmeras, me incentivar a aprender e estar sempre estudando, além de me aventurar em análises e métodos novos. Obrigada por sempre me desafiar como profissional, por me tornar mais questionadora e independente. Levo comigo todos os seus conselhos e, principalmente, o exemplo de profissional e orientador dedicado e determinado em melhorar a situação epidemiológica do país. Muito obrigada mesmo!

Ao Dr. Amilcar Tanuri pela confiança depositada em mim ao disponibilizar as sequências e dados para que esse trabalho fosse desenvolvido, bem como pela oportunidade de desenvolver parte das análises no Laboratório de Virologia Molecular/UFRJ e pela acolhida durante minha estadia.

Aos pesquisadores e alunos do Laboratório de Virologia Molecular/UFRJ por todo carinho na acolhida durante meus períodos de trabalho e disciplinas no Rio de Janeiro, por toda as conversas, risadas e almoços no refeitório, troca de experiência, companhia no congresso de virologia e patogênese, além do auxílio nas análises ao longo desse trabalho. Certamente ficar longe de casa foi mais fácil com vocês ao lado. Agradecimento especial a Camila, Lucas, Filipe, Mariana, Pedro, Luiza, Alice, Átila e Fernando por toda atenção, carinho e força nessa caminhada. A Paula do PPG-BCS/FiocruzRJ por ter compartilhado comigo esses momentos de desafios na bioinformática, que nosso esforço não tenha sido em vão. Obrigada a todos pela amizade.

Aos meus colegas de LIA, por toda a convivência, troca de experiências, discussão sobre as análises e seminários. Muito obrigada por todas as considerações feitas ao longo do trabalho.

Aos colegas de pós-graduação Sayonara, Aline, Lander, Leandra, Caibe, Laryssa, Tatiane, Gabriel e alunos do laboratório de protozoologia por toda amizade, risadas, roda de mate nas disciplinas, conversas e momentos de desespero compartilhados ao longo do mestrado. Em especial ao Guilherme por me auxiliar desde a seleção até todos os socorros em Linux no decorrer das análises. Muito obrigada.

Aos funcionários do LAMEB Denis, Vanessa, Maísa, William e Izabella do LVA por toda a amizade, apoio, cafés roubados e momentos de lazer fora da UFSC. Obrigada por me aceitarem como agregada e me adotarem como família durante o mestrado.

Aos professores do programa de pós-graduação em Biotecnologia e Biociências, em especial aos professores Dra. Patrícia Stoco, Dr. Edmundo Grisard, Dr. Glauber Wagner, Dr. Rafael Rosa, Dr. José Henrique Oliveira por despertaram o em mim o espírito questionador. Professores e pesquisadores que exercem a ciência de forma crítica e com qualidade.

RESUMO

A epidemia de HIV-1 no sul do Brasil é predominantemente causada pelo HIV-1C, com co-circulação de subtipos B, F1 e formas recombinantes. O estado de Santa Catarina apresenta elevadas taxas de detecção, mortalidade e incidência, sendo Florianópolis a capital com o pior cenário epidemiológico. Os fatores que contribuem para as contrastantes diferenças observadas na epidemiologia de HIV/aids nessa região ainda não são bem compreendidos, contudo podem estar relacionadas ao comportamento de risco diferenciado desta população ou pela diversidade molecular única encontrada nessa região. Embora diversos estudos tenham procurado descrever a diversidade molecular, os dados coletados focaram nos maiores centros afetados pelo HIV, não descrevendo a população do estado como um todo. Esse trabalho tem por objetivo descrever a diversidade genética do HIV-1 em Santa Catarina em amostras coletadas de pacientes em falha terapêutica, detalhando variações de prevalência no tempo e espaço territorial, além de associações com dados epidemiológicos, tentando elucidar as redes de transmissão intermunicipais, interestaduais e internacionais envolvidas na disseminação da epidemia utilizando-se, para isso, abordagem filogenética. O HIV-1C mostrou-se como a forma viral mais relevante na epidemia em Santa Catarina, com significativo aumento de prevalência ao longo dos anos, se sobrepondo a formas virais mais raras. Esse estudo indicou também que HIV-1C e B circulam em populações epidemiologicamente distintas, sendo a infecção por subtipo C mais associada ao gênero feminino e indivíduos com baixa escolaridade, enquanto o B mostrou-se mais presente em homens e indivíduos com maior grau educacional. O aumento evidenciado para esse subtipo pode estar associado a introdução do HIV-1C em grupos HSH (homens que fazem sexo com homens), visto que, embora tenha-se evidenciado forte associação entre HIV-1B e indivíduos possivelmente HSH, o subtipo C foi encontrado em alta proporção nesses indivíduos, indicando uma intercomunicação das redes de transmissão. Foi observado também uma distribuição heterogênea de subtipos pelas regiões intermediárias do estado, podendo esse fato estar associado as redes de disseminação virais, fluxos migratórios, regiões de fronteira e turismo. O subtipo C apresentou as maiores taxas de clusterização quando comparado aos outros subtipos, além de maior número, tamanho e distância entre os membros contidos nos agrupamentos identificados, sugerindo que as redes de transmissão onde esse subtipo circula é mais abrangente e a epidemia mais disseminada. Somado a isso, esse subtipo apresentou maior relevância em interações epidemiológicas internacionais e interestaduais, indicando que as redes de transmissão atuantes em Santa Catarina podem contribuir para a disseminação de subtipo C para outras localidades. Em conclusão, esse estudo corrobora a importância epidemiológica do subtipo C para o estado de Santa Catarina, bem como indica que as redes de transmissão atuantes no estado podem contribuir para o potencial expansivo dessa epidemia, principalmente, em outras regiões do país.

Palavras-chave: HIV, Subtipo C, Epidemiologia Molecular, Redes de Transmissão, Dispersão Viral.

ABSTRACT

The HIV-1 epidemic in southern Brazil is mostly caused by HIV-1C, with co-circulation of subtype B, F1 and recombinant forms. The Santa Catarina state presents high detection, mortality and incidence rates, with Florianópolis being the national capital with the worst epidemiological scenario. The factors that trigger the contrasting differences observed in the epidemiology of HIV /AIDS in this region are not well understood yet, and may be related to the differentiated risk behavior of the population or the unique HIV molecular diversity. Although other studies have aimed to describe the molecular diversity, data collection has focused on the major cities affected by HIV, not describing the state population as a whole. This work aims to describe the genetic diversity of HIV-1 in the state of Santa Catarina, detailing the variations in time and space and associations with demographical features, as well as trying to elucidate the intrastate, interstate and external transmission networks involved in the spread of the epidemic using phylogenetic approach. HIV-1C proved to be the most relevant viral form in the Santa Catarina epidemic, with a significant increase in prevalence over the years overlapping with rarer viral forms. This study also indicated that HIV-1C and B circulate in epidemiologically distinct populations, with subtype C infection being more associated with females and individuals with low education, while B was more present in men and individuals with higher educational level. The increase seen for this subtype may be associated with the introduction of HIV-1C in MSM groups, although although there was a strong association between HIV-1B and possibly MSM individuals, subtype C was found in high proportion in these individuals, indicating an intercommunication of the transmission networks. It was also observed a heterogeneous distribution of subtypes across the intermediate regions of the state, which may be associated with viral dissemination networks, migratory flows, border regions and tourism. The subtype C presented the highest clustering rates when compared to the other subtypes, besides the larger number, size and distance between the limbs contained in the identified clusters, suggesting that the transmission networks where this subtype circulates is broader and the epidemic more widespread. In addition, this subtype was more relevant in international and interstate epidemiological interactions, indicating that the transmission networks operating in Santa Catarina may contribute to the dissemination of subtype C to other locations. In conclusion, this study corroborates the epidemiological importance of subtype C for the state of Santa Catarina, as well as indicates that the transmission networks operating in the state may contribute to the expansive potential of this epidemic, especially in other regions of the country.

Key words: HIV, Aids, Subtype C, Molecular Epidemiology, Phylodynamics, Viral Dispersion, Brazilian Southern Region.

Lista de Figuras

Figura 1: Representação esquemática do mapa genômico do HIV-1, obtido a partir do Los Alamos HIV Database (2019).	20
Figura 2: Representação gráfica da história natural da infecção pelo HIV-1. Adaptado de Goulder et al., 2004 e obtida de Gräf, 2015.	21
Figura 3: Distribuição mundial dos subtipos do grupo M do HIV-1 e formas recombinantes circulantes (CRFs). Adaptado de Ariën et al., 2007 e obtida de Gräf, 2015.	26
Figura 4: Representação da diversidade molecular do HIV-1 em cidades do interior do Rio Grande do Sul e Santa Catarina. Adaptado de Gräf, 2015.	29
Figura 5: Representação esquemática de uma reconstrução filogenética e determinação de possíveis relações de transmissão. Adaptado de Dennis <i>et al.</i> , 2014.	33
Figura 6: Representação da influência da determinação de ponto de corte de distância genética na quantidade e tamanho dos <i>clusters</i> de transmissão identificados. Adaptado de Hassan <i>et al.</i> , 2017.	38
Figura 7: Fluxograma de metodologias aplicadas para subtipagem das sequências da RENAGENO-SC.....	51
Figura 8: Variação anual da frequência dos subtipos de maior relevância epidemiológica descritos em pacientes com falha terapêutica no estado de Santa Catarina.....	54
Figura 9: Regressão linear de frequência do subtipo no ano (eixo y) versus data de coleta (eixo x).	55
Figura 10: Diferenças na segregação espacial do HIV-1 B, C, F1 e CRF31_BC nas regiões intermediárias do estado de Santa Catarina.....	60
Figura 11: Fluxograma de abordagens utilizadas na seleção de sequências externas ao dataset RENAGENO-SC.	71
Figura 12: Correlação entre o número de sequências amostradas por ano e o número de sequências agrupadas no <i>dataset</i> RENAGENO-SC.....	78

Figura 13: Distribuição da frequência obtida e tamanhos de <i>clusters</i> de transmissão para os subtipos B, C, F1 e CRF31_BC do HIV-1.	81
Figura 14: Representação filogenética do maior <i>cluster</i> de transmissão do HIV-1C, identificado pela ferramenta Cluster Picker, no estado de Santa Catarina. Amostragem de 10 indivíduos potencialmente relacionados epidemiologicamente por eventos de transmissão.	82
Figura 15: Distribuição das distâncias médias entre indivíduos compondo <i>clusters</i> de transmissão exclusivamente RENAGENO-SC.	90

Lista de Tabelas

Tabela 1: Dados demográficos referentes as quatro principais formas de HIV-1 encontradas no banco de dados RENAGENO-SC.....	57
Tabela 2: Dados epidemiológicos associados com o agrupamento geral em pacientes da RENAGENO-SC.	77
Tabela 3: Análise comparativa das redes de transmissão identificadas para os subtipos B, C, F1 e CRF31_BC do HIV-1.	80
Tabela 4: Dados epidemiológicos associados as diferenças em agrupamentos observados para cada subtipo avaliado.	84
Tabela 5: Caracterização da composição geográfica das redes de transmissão para os subtipos B, C, F1 e CRF31_BC.	87
Tabela 6: Caracterização de tamanho e distância entre membros de <i>clusters</i> de transmissão constituídos exclusivamente por pacientes RENAGENO-SC.	89
Tabela 7: Agrupamento diferencial de crianças com idade inferior a 14 anos e associação com dados epidemiológicos.	92
Tabela 8: Características epidemiológicas associadas às possíveis mães agrupadas com crianças com idade inferior a 14 anos.	94
Tabela 9: Características epidemiológicas associadas a indivíduos do sexo masculino que clusterizaram exclusivamente com homens.	96
Tabela 10: Comparação de tamanho de clados obtidos com Cluster Picker, a partir de árvore com e sem TBE.....	126
Tabela 11: Comparação do tamanho dos <i>clusters</i> obtidos para os subtipos B, C, F1 e CRF31_BC a partir de árvores com e sem TBE.....	127

Lista de Abreviaturas e Siglas

aLRT: teste de razão de verossimilhança aproximada

ARV: antirretroviral

BI: Inferência Bayesiana

CCR5: receptor de quimiocina C-C tipo 5

CD4: grupamento de diferenciação 4 positivo

cluster: agrupamento

CRF: forma recombinante circulante

CRF31_BC: forma recombinante circulante 31 BC

CXCR4: receptor de quimiocina C-X-C tipo 4

DNA: ácido desoxirribonucléico

env: gene do envelope

EUA: Estados Unidos da América

gag: gene antígeno grupo-específico

gp120: glicoproteína de superfície 120

gp41: glicoproteína de superfície 41

HET: heterossexual

HIV: vírus da imunodeficiência humana

HIV-1: vírus da imunodeficiência humana tipo 1

HIV-1B: subtipo B do HIV-1

HIV-1C: subtipo C do HIV-1

HIV-1F1: subtipo F1 do HIV-1

HSH: homem que faz sexo com homem

IBGE: Instituto Brasileiro de Geografia e Estatística

INTI: inibidores de integrase

ML: Máxima Verossimilhança

nef: fator de regulação negativa

NJ: Neighbor-Joining

NNRTI: inibidores não-nucleosídicos da transcriptase reversa

NRTI: inibidores nucleosídicos da transcriptase reversa

Pb: pares de base

PI: inibidores da protease

pol: gene da polimerase

PP: probabilidade posterior
PR: Paraná
RENAGENO: Rede Nacional de Genotipagem
RNA: ácido ribonucléico
RS: Rio Grande do Sul
SC: Santa Catarina
SH-like: teste Shimodaira-Hasegawa
TARV: terapia antirretroviral
Tat: proteína viral transativadora de transcrição
TBE: *transfer bootstrap expectation*
TV: transmissão vertical
UDI: usuários de drogas injetáveis
URF: forma recombinante única
rev: regulador de expressão de proteínas virais
vif: fator de infectividade viral
vpr: proteína viral R
vpu: proteína viral U

SUMÁRIO

1. INTRODUÇÃO.....	19
1.1 O vírus da imunodeficiência humana (HIV) e a aids	19
1.2 Epidemiologia molecular do HIV-1	24
1.2.1 Epidemiologia molecular do HIV-1 no Brasil e situação peculiar da região Sul	26
.....	29
1.2.2 Situação epidemiológica do estado de Santa Catarina	29
1.3 Redes de transmissão viral	30
1.4 Ferramentas para detecção de redes de transmissão.....	33
1.4.1 Influência da determinação de ponto de corte para suportes de ramo e distância genética	36
1.4.2 Influência da densidade amostral e uso para monitoramento de grandes epidemias e epidemias concentradas	41
1.4.3 Influência do tamanho e escolha de fragmento gênico.....	42
2. OBJETIVOS.....	45
OBJETIVO GERAL	45
OBJETIVOS ESPECÍFICOS	45
3. O BANCO DE DADOS E ASPECTOS ÉTICOS	46
CAPÍTULO 1 – Análise temporo-espacial da diversidade genética e epidemiologia molecular do HIV-1 em pacientes em falha terapêutica do estado de Santa Catarina, região sul do Brasil.	48
4. METODOLOGIA	49
4.1 Controle de qualidade e subtipagem das sequências RENAGENOS-SC.....	49
4.2 Análises Estatísticas	52
5. RESULTADOS	53

5.1	Caracterização molecular da epidemia de HIV-1 no estado de Santa Catarina.....	53
5.2	Características epidemiológicas dos pacientes da RENAGENO-SC.....	55
5.3	Dispersão heterogênea de subtipos de HIV-1 nas regiões intermediária de Santa Catarina	58
6.	DISCUSSÃO.....	61
	CAPÍTULO 2 – Análise da dinâmica de transmissão do HIV-1 em pacientes em falha terapêutica do estado de Santa Catarina, região sul do Brasil.	68
7.	METODOLOGIA	69
7.1	Remoção de sítios associados a mutações de resistência aos ARV	69
7.2	Seleção de sequências externas brasileiras e internacionais	69
7.3	Identificação das redes de transmissão	72
7.4	Caracterização das redes de transmissão	73
7.5	Análises Estatísticas	74
8.	RESULTADOS	75
8.1	Análise comparativa das características epidemiológicas associadas ao agrupamento.....	75
8.2	Análise comparativa das características epidemiológicas associadas à clusterização nos diferentes subtipos	79
8.3	Análise da composição geográfica das redes de transmissão e agrupamento externo.....	85
8.4	Análise geoespacial da distribuição dos <i>clusters</i> de transmissão em Santa Catarina.....	88
8.5	Análise de agrupamento em indivíduos com idade inferior a 14 anos e indicativo de possível transmissão vertical	91
8.6	Análise das redes de transmissão formadas exclusivamente por indivíduos do sexo masculino	95

9. DISCUSSÃO.....	98
10. SUMÁRIO DE RESULTADOS.....	107
11. CONCLUSÃO	108
12. PERSPECTIVAS	109
REFERÊNCIAS BIBLIOGRÁFICAS	110
APÊNDICES.....	126

1. INTRODUÇÃO

1.1 O vírus da imunodeficiência humana (HIV) e a aids

O HIV é um vírus pertencente à família *Retroviridae*, sendo capaz de sintetizar, durante seu ciclo replicativo, um genoma intermediário de DNA a partir de uma molécula molde de RNA. Após a infecção apresenta um longo período de incubação e lento desenvolvimento da sintomatologia, semelhante ao observado em outros vírus do gênero *Lentivirus* (Hoffmann et al., 2007). Existem dois tipos de HIV: HIV-1 e HIV-2. Ambos apresentam morfologia semelhante, contudo apresentam diferenças marcantes quanto à patogenicidade e dispersão mundial. O HIV-1 é conhecido por ser o agente causador da pandemia mundial de aids enquanto o HIV-2 mostra-se menos disperso, sendo relatado como causador de infecções restritas ao oeste da África e algumas localidades da Europa (Rambaut et al., 2004).

O genoma do vírus é composto por duas fitas simples e idênticas de RNA, contendo aproximadamente 10 mil nucleotídeos cada, que codificam nove genes (Figura 1). Destes, três são vitais para o processo infeccioso e encontram-se presentes em todos os retrovírus: *env*, *gag* e *pol*. O gene *env* codifica para glicoproteínas de superfície (gp120 e gp41), que compõe o envelope viral; o gene *gag* é responsável pela codificação de proteínas estruturais constituintes do capsídeo; o gene *pol* codifica as enzimas virais transcriptase reversa, protease e integrase (Hoffmann et al., 2007). Os outros seis genes denominados *vif*, *vpu*, *vpr*, *tat*, *rev* e *nef* desempenham funções acessórias ao mecanismo de infecção do HIV-1 (Miller e Sarver, 1997; Cullen, 1998).

O ciclo de replicação do vírus inicia com a interação das glicoproteínas do envelope viral com o receptor CD4 presente na superfície de células do sistema imunológico tais como linfócitos T, monócitos, macrófagos, células dendríticas e micróglia. Esta primeira interação não é suficiente para que ocorra a fusão das membranas da célula hospedeira e viral, necessitando ser estabilizada mediante interação com co-receptores, sendo os mais utilizados CCR5 ou CXCR4. Após a entrada do vírus na célula, o capsídeo é

desintegrado, liberando o material genético viral no citoplasma. Depois disso, as enzimas virais iniciam o processo de transcrição reversa da fita de RNA, sendo que o DNA dupla fita é translocado até o núcleo e integrado no genoma celular. Após o evento de integração, o segmento de DNA viral é transcrito como uma poliproteína clivada pela enzima protease, possibilitando que as proteínas virais se organizem adequadamente na montagem do vírion, para posterior brotamento (Hoffmann et al., 2007).

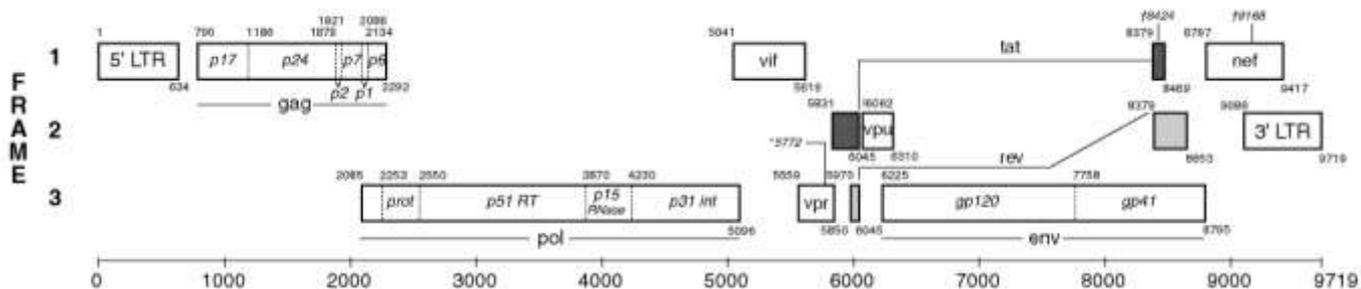


Figura 1: Representação esquemática do mapa genômico do HIV-1, obtido a partir do Los Alamos HIV Database (2019).

Representação do posicionamento dos segmentos gênicos compondo o genoma do HIV-1 e sua distribuição segundo as fases de leitura de leitura. O início do gene encontra-se indicado pelo pequeno número no canto superior esquerdo de cada retângulo, normalmente registrando a primeira posição do códon de início ATG para esse gene, enquanto o número no canto inferior direito registra a última posição do códon de parada.

A principal célula-alvo da infecção viral por HIV-1 são linfócitos T CD4. Este tipo celular sofre depleção ao longo da evolução da doença, sendo sua redução associada ao quadro de imunodeficiência e acometimento do paciente por doenças oportunistas. A história natural da doença (Figura 2) pode ser dividida em três fases: aguda (logo após a infecção), crônica e fase de aids. Na fase aguda o indivíduo apresenta elevada viremia, entretanto os sintomas, quando acontecem, são inespecíficos e semelhantes a um quadro gripal. Após algumas semanas, a carga viral circulante é reduzida devido a ação de linfócitos T citotóxicos HIV-específicos (Mellors et al., 1995).

A fase crônica da infecção tem como principal característica um período de equilíbrio entre a replicação viral e resposta imune do hospedeiro frente ao

processo infeccioso. Esta fase é assintomática e pode perdurar anos; contudo, há progressiva depleção de linfócitos T CD4. Com a redução destas células, o indivíduo passa a exibir um perfil de imunossupressão, tornando-se mais susceptível a infecções oportunistas. A fase de aids é atingida quando a contagem de linfócitos T CD4 encontra-se inferior a 350 células/ μL de sangue associada ao aparecimento de infecções oportunistas fúngicas e bacterianas ou neoplasia, tal como sarcoma de Kaposi (Ministério da Saúde, Coordenação de HIV/aids, 2018). Neste momento o paciente apresenta elevada viremia devido à ineficiência do sistema imunológico em controlar a replicação viral, bem como grave imunossupressão e deficiência em combater infecções por patógenos oportunistas (Hoffmann et al., 2007).

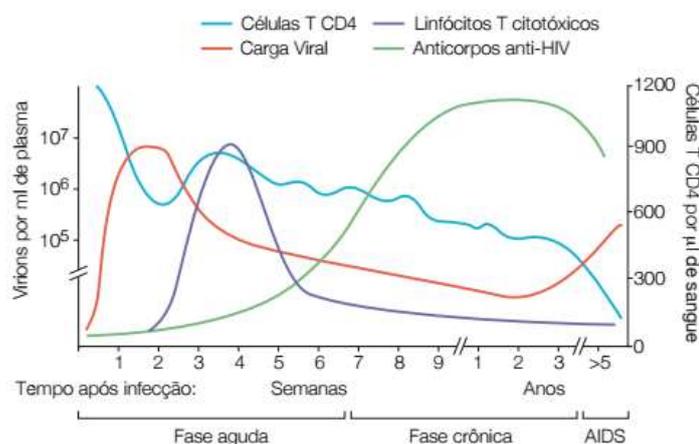


Figura 2: Representação gráfica da história natural da infecção pelo HIV-1. Adaptado de Goulder et al., 2004 e obtida de Gräf, 2015.

A linha azul claro representa o comportamento de declínio em número das células T CD4 e em roxo os linfócitos T citotóxicos, abundantes durante a fase aguda de infecção. A linha vermelha representa o comportamento da carga viral ao longo da infecção, apresentando um aumento drástico durante o início da fase aguda, redução e estabilização durante o final da fase aguda e fase crônica, seguido de um aumento acentuado durante o final da fase crônica e início da aids. Em verde é representado o aumento na detecção de anticorpos anti-HIV durante o curso clínico, sendo mais abundantes e de fácil detecção durante a fase crônica.

Com avanços no desenvolvimento de novos medicamentos e o manejo de diferentes vias para controle da infecção viral pelo HIV, a implementação da TARV (terapia antirretroviral) propiciou um aumento progressivo nas taxas de

resposta terapêutica. A partir de 2014, o Ministério da Saúde brasileiro preconizou que todo paciente diagnosticado com HIV inicie terapia de forma imediata (Ministério da Saúde, 2017), alternando o antigo regime de início de tratamento após o indivíduo atingir contagem de células T CD4+ inferior a 200 células/mm³ de sangue. Com os esquemas antirretrovirais modernos, cerca de 80% dos pacientes passam a apresentar carga viral indetectável (inferior a 50 cópias/mL) após o primeiro ano de tratamento, sendo a maioria capaz de manter essa supressão viral medicamentosa pelos anos seguintes. Contudo, ainda há uma parcela dos pacientes em tratamento que apresentam falha ao tratamento inicial, sendo necessário a aplicação de uma TARV de resgate (Ministério da Saúde, Coordenação de HIV/AIDS, 2018). Atualmente encontram-se aprovados para utilização em terapia antiretroviral 25 medicamentos divididos em 6 classes distintas (Broder, 2010):

- i. Inibidores nucleosídicos da transcriptase reversa (NRTI): o mecanismo de ação baseia-se na atuação como falso substrato para a enzima RT, competindo com os nucleotídeos da célula infectada, ocorrendo um impedimento na síntese de DNA pró-viral;
- ii. Inibidores não-nucleosídicos da transcriptase reversa (NNRTI): atuam ligando-se próximo ao sítio catalítico da enzima RT, retardando a incorporação de novos nucleotídeos na fita de DNA em elongação;
- iii. Inibidores da protease (PI): atuam inibindo a ação enzimática da protease viral, impedindo a formação de proteínas constituintes do capsídeo viral e outras enzimas do HIV;
- iv. Inibidores de fusão: o mecanismo de ação é baseado na ligação dele com sequências peptídicas de caráter hidrofóbico presentes na gp41 responsáveis pela fusão do envelope viral com a membrana plasmática da célula hospedeira, impedindo a entrada do vírus;
- v. Inibidores de entrada: essa classe de antirretroviral é composta por moléculas sintéticas que apresentam efeito de antagonista ao

receptor de quimiocinas CCR5, impedindo que ocorra correta interação da gp120 viral com o co-receptor de auxílio;

- vi. Inibidores de integrase (INTI): ligam-se à enzima integrase, impedindo que ocorra o processo de incorporação do DNA pró-viral no DNA genômico da célula hospedeira.

O resgate no controle da viremia de pacientes refratários ao esquema medicamentoso é dependente do reconhecimento precoce da falha virológica e emprego cuidadoso de estratégias para correto delineamento do novo tratamento. A falha virológica pode ser caracterizada por carga viral detectável após seis meses do início ou modificação de tratamento, como também por aumento inesperado na carga viral de pacientes que já haviam atingido a supressão viral sob tratamento. Esse processo de aumento da carga viral prejudica a recuperação imunológica dos níveis de células T CD4+, além de aumentar o risco de progressão da doença e emergência de cepas resistentes aos ARV. A supressão parcial e a persistência dos níveis detectáveis de vírus circulante facilitam não só o acúmulo de mutações que conferem resistência frente à medicação em uso, como também reduz, a longo prazo, as opções terapêuticas (Ministério da Saúde, Coordenação de HIV/AIDS, 2018).

Uma vez detectada e confirmada a falha virológica a partir de duas coletas subsequentes de sangue (com intervalo de quatro semanas entre elas), recomenda-se a genotipagem da população viral e pesquisa de resistência aos ARV, auxiliando assim a elaboração de um esquema de resgate mais efetivo na supressão viral. No Brasil, o teste de genotipagem de pacientes em falha terapêutica é oferecido no Sistema Único de Saúde (SUS), através da Rede Nacional de Genotipagem (RENAGENO). Essa testagem consiste em sequenciar parte do genoma viral e pesquisar, em segmentos gênicos específicos (gene *pol*), a presença de códons associados à resistência às diferentes drogas utilizadas no tratamento. Com isso é possível otimizar a escolha de esquemas TARV com maior chance de sucesso, propiciar o uso de medicamentos ativos por períodos prolongados, prevenir a troca desnecessária de esquema terapêutico, além de reduzir a toxicidade por medicamentos

inativos e melhorar a relação custo-efetividade do tratamento (Ministério da Saúde, 2018).

1.2 Epidemiologia molecular do HIV-1

A principal característica do HIV-1, que resulta na alta diversidade molecular apresentada por esse patógeno e no grande dinamismo da epidemia de aids, é sua elevada taxa evolutiva (Tebit e Arts, 2011). A diversidade molecular encontrada nesse vírus é decorrente de uma combinação de fatores que o torna um dos organismos que apresentam evolução mais rápida e passível de mensuração em uma curta escala de tempo (Rambaut et al., 2004; Grenfell et al., 2004). Essa rápida evolução é associada à inexistência de processos de correção de erros da transcriptase reversa, ocasionando alta taxa mutacional a cada processo de transcrição (0,2 mutações por genoma) (Preston et al., 1988). Ademais, esse acúmulo de mutações torna-se ainda mais acentuado devido ao número de 10^{10} a 10^{12} novos vírions (partículas virais) produzidos por dia (Perelson et al., 1996). Com isso, o HIV-1 possui uma taxa evolutiva estimada em 2 a 4×10^{-3} substituições por sítio a cada ano (Abecasis et al., 2009). Desta forma, considerando o genoma de aproximadamente 10 mil pares de bases do HIV-1, cerca de 20 a 40 mutações são fixadas em uma específica população viral, o que equivale a 0,2 a 0,4% de divergência genética acumulada por ano.

Em decorrência dessa rápida taxa evolutiva do HIV-1, a capacidade de resposta imunológica do organismo frente ao vírus torna-se deficitária. Além disso, a ampla diversidade de variantes virais inviabiliza o desenvolvimento de vacinas capazes de atuar com elevada eficácia e com amplo espectro de ação (Ramos et al., 1999), bem como possibilita mecanismos de evasão e resistência aos tratamentos disponíveis, ocasionando falha terapêutica (Grenfell et al., 2004).

A atuação de diferentes mecanismos evolutivos intrínsecos ao vírus e dependentes do hospedeiro possibilitou que o HIV-1 pudesse divergir amplamente em seu genoma, sendo classificado em quatro grupos: M, N, O e P. Os grupos N e O apresentam distribuição discreta, sendo endêmicos em regiões do oeste africano. O grupo P é descrito como um outro agrupamento

do HIV-1 com baixa prevalência, tendo sido isolado de uma paciente de Camarões (Plantier et al., 2009). O grupo M, entretanto, apresenta maior relevância epidemiológica e clínica, visto que é o responsável pela pandemia atual da doença. Esse grupo é subdividido em subtipos puros (A-D, F-H, J e K) (Robertson et al, 2000; Rambaut et al., 2004), além de alguns sub-subtipos (A1-A4, A6 e F1-F2) (Los Alamos HIV Database, 2018). A elevada diversidade viral observada no grupo M também compreende 95 formas recombinantes intersubtipos (Los Alamos HIV Database, 2019), que até o ano de 2012-2013 correspondiam a 20% das novas infecções mundiais (Lau e Wong, 2013), além de infecções causadas por formas recombinantes únicas (URFs).

Segundo a UNAIDS atualmente há mais de 37,9 milhões de pessoas vivendo com HIV-1 no mundo (WHO/UNAIDS, 2019). Os subtipos virais associados às infecções e os números de indivíduos acometidos distribuem-se heterogeneamente no território global (WHO/UNAIDS, 2019), sendo a África Central a região que apresenta a epidemiologia molecular mais diversificada (Hemelaar et al., 2006; Osmanov et al., 2002). O subtipo C é o mais prevalente mundialmente, sendo encontrado na Índia e região sul da África, compreendendo aproximadamente 50% das infecções globais. Já o subtipo B é a forma mais dispersa no mundo, sendo o principal responsável pela maioria de casos de aids no continente americano, Europa e Oceania (Ariën et al., 2007) (Figura 3). Estudos indicam que esse subtipo do HIV-1 seja responsável por aproximadamente 12% das infecções ao redor do mundo; contudo, embora menos prevalente que o subtipo C, a maioria das pesquisas clínicas sobre HIV-1 têm sido conduzidas em populações onde esse subtipo é predominante.

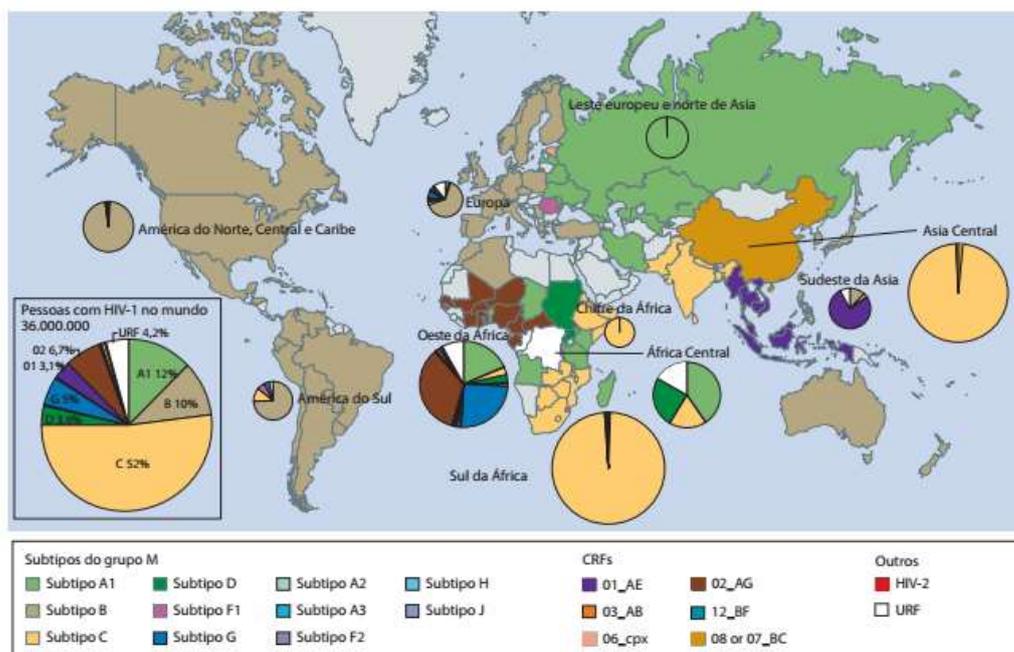


Figura 3: Distribuição mundial dos subtipos do grupo M do HIV-1 e formas recombinantes circulantes (CRFs). Adaptado de Ariën et al., 2007 e obtida de Gräf, 2015.

É possível observar que o subtipo C é o mais prevalente no mundo, sendo responsável por 52% dos indivíduos infectados e sendo majoritário em países populosos, enquanto o subtipo B encontra-se causando aproximadamente 10% das infecções, sendo a forma viral responsável por infecções no continente americano e europeu.

1.2.1 Epidemiologia molecular do HIV-1 no Brasil e situação peculiar da região Sul

O Brasil é um país de dimensões continentais. Essa característica propicia que ocorram variações epidemiológicas acerca da epidemia de HIV/aids conforme a região geográfica analisada. Segundo o Ministério da Saúde, desde 1980 até os dias atuais foram notificados aproximadamente 966 mil indivíduos vivendo com HIV-1 no Brasil (Ministério da Saúde, 2019). A maior concentração de casos de aids é observada em indivíduos entre 25 e 39 anos, em ambos os sexos (Ministério da Saúde, 2019), sendo que a prevalência de aids entre indivíduos de 15 a 49 anos permanece estável em torno de 0,6% desde 2004 (Ministério da Saúde, 2017). A taxa de detecção anual de novos casos está apresentando queda desde 2013. A incidência até

então era estável em torno de 21 novos casos a cada 100 mil habitantes, passando para 17,8 em 2019, devido às modificações no protocolo de tratamento e notificação do Ministério da Saúde ocorrido em 2014.

Na região Sul, apesar da taxa de detecção de novos casos de aids também estar em queda, é muito mais alta que a média nacional, tendo passado de 35,7 casos a cada 100 mil habitantes em 2008, para 22,8 em 2018. Devido a essa elevada incidência de aids na região, a taxa de mortalidade também é elevada (5,5 a cada 100 mil habitantes), sendo a segunda maior em comparação ao restante do país (Ministério da Saúde, 2019). Entre os estados desta região, Rio Grande do Sul e Santa Catarina apresentam os piores cenários, sendo que em 2018 a taxa de detecção de aids foi de 27,2 e 25,6 casos a cada 100 mil habitantes, respectivamente. Além disso, Florianópolis é a cidade que apresentou a maior incidência entre as capitais do país, com taxas de detecção de 57 casos para cada 100 mil habitantes, número quase três vezes maior que a taxa brasileira (Ministério da Saúde, 2019). Além disso, Florianópolis apresenta um coeficiente de mortalidade duas vezes maior que o restante do país (8,5 por 100 mil habitantes) (Ministério da Saúde, 2019).

Outra característica peculiar da epidemia no sul do Brasil é a predominância do HIV-1 subtipo C. Enquanto no Brasil como um todo o subtipo B é responsável por 70-90% das infecções (Teixeira et al., 2004; Sa Filho et al., 2005; Gadelha et al., 2003; Cerqueira et al., 2004; Couto-Fernandez et al., 2006; Pereira et al., 2006), a circulação do subtipo C não ultrapassa 3% (Pedroso et al., 2007; Cardoso et al., 2009, Sa Filho et al., 2005; Gadelha et al., 2003; Machado et al., 2009). Contudo, no estado do Rio Grande do Sul o HIV-1C é encontrado com prevalência variando entre 40 e 56% nas principais cidades (Simon et al., 2010, de Medeiros et al., 2011, Silveira et al., 2012). No estado de Santa Catarina, diversos estudos transversais conduzidos observaram um aumento progressivo no número de infecções ocasionadas por subtipo C, com prevalências variando entre 48% a valores superiores a 60% (Locateli et al., 2007, Brígido et al., 2007; Rodrigues et al., 2010; Gräf et al., 2011, Gräf et al., 2016). Ademais, também foram relatadas infecções por formas mosaicos BC (aproximadamente 10%), evidenciando a co-existência de ambos subtipos (B e C) na mesma população (Gräf et al., 2016). Embora haja

fortes indícios de um processo de expansão de epidemia de HIV-1C no estado catarinense, nenhum estudo longitudinal foi conduzido a fim de caracterizar o crescimento e a dispersão dessa forma viral no território, bem como discriminar fatores epidemiológicos que podem ser atuantes nesse fenômeno.

Em contramão ao observado no restante da região Sul, o estado do Paraná apresenta frequência de subtipo C aproximadamente de 2 a 3 vezes menor ao observado nos outros estados (Brindeiro et al., 2003, Ferreira et al., 2008, Toledo et al., 2010; Raboni et al., 2010; Silva et al., 2010). Em vista dos valores de prevalência de HIV-1C entre Paraná e os outros estados do Sul serem discrepantes, torna-se vital avaliar possíveis interconexões entre esses estados e determinar fatores epidemiológicos que podem, futuramente, ocasionar um aumento na dispersão dessa forma viral para outros estados e regiões.

Além da elevada prevalência do subtipo C, a epidemia de HIV-1 no Sul também é caracterizada pela circulação de formas recombinantes BC, dentre elas a CRF31_BC (Guimaraes et al., 2002; Soares et al., 2003; Brindeiro et al., 2003; Santos et al., 2006; Brígido et al., 2007). Estudos recentes evidenciaram também a dispersão da forma recombinante CRF31_BC para o estado de Santa Catarina através de transições envolvendo principalmente a cidade de Porto Alegre (Fritsch et al., 2018). Em estudo conduzido por Gräf et al (2015), Porto Alegre é descrita como ponto geográfico vital para disseminação de subtipo C para as regiões sudeste, norte e centro-oeste do país, destacando-se transições com alto suporte para diferentes cidades de Santa Catarina (Figura 4). Somado a isso, Raboni et al (2010) observou maior proporção de subtipo C (31%) em cidades próximas à fronteira com Santa Catarina, enquanto no norte do estado a prevalência foi três vezes menor (13%). Tais dados corroboram a hipótese de que há uma alta interconexões entre as redes de transmissão de Santa Catarina e estados vizinhos.

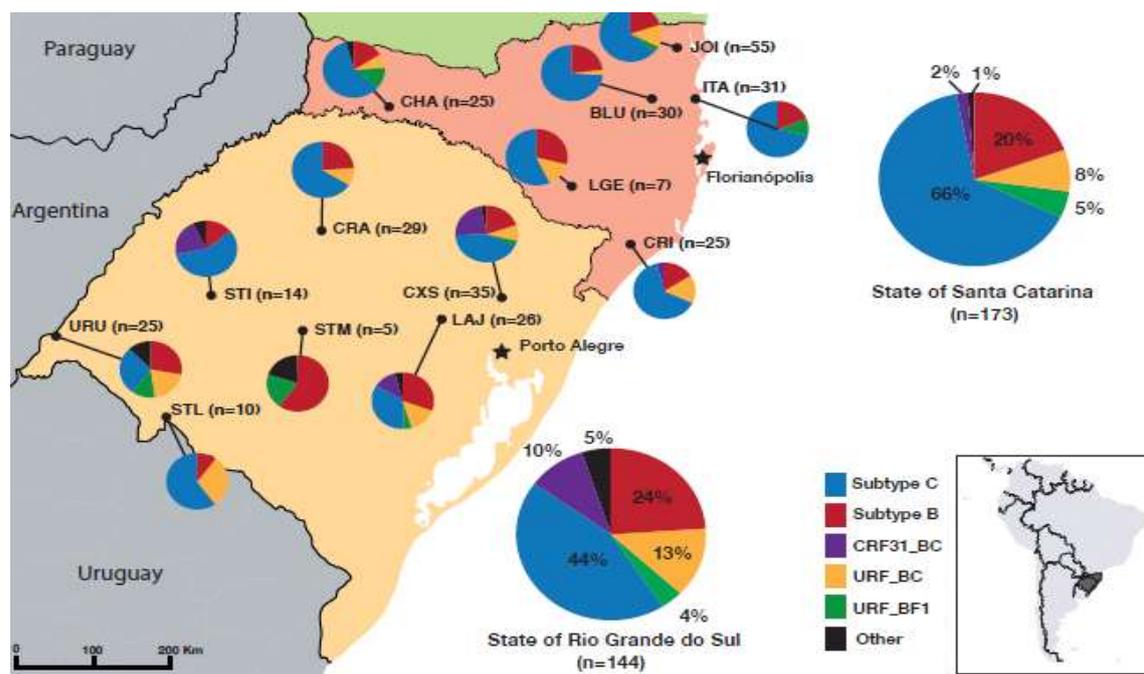


Figura 4: Representação da diversidade molecular do HIV-1 em cidades do interior do Rio Grande do Sul e Santa Catarina. Adaptado de Gräf, 2015.

URU: Uruguaiana, STI: Santiago, STM: Santa Maria, CXS: Caxias do Sul, LAJ: Lajeado, CRA: Cruz Alta, STL: Santana do Livramento, CRI: Criciúma, LGE: Lages, CHA: Chapecó, BLU: Blumenau, JOI: Joinville, ITA: Itajaí

1.2.2 Situação epidemiológica do estado de Santa Catarina

O estado de Santa Catarina (SC) encontra-se localizado na região Sul do país e possui aproximadamente 7 milhões de habitantes (IBGE, 2017). O primeiro registro de aids ocorreu na região oeste do estado em 1984 e até o ano de 2019 foram notificados 49.585 casos de aids, afetando 71,3% dos municípios do estado até o ano de 2004 (D'orsi, 2004 *apud* Locatelli et al., 2007; Ministério da Saúde, 2019). Desde o primeiro caso notificado, o estado contabiliza um total de mais de 12 mil óbitos decorrentes da infecção pelo HIV-1. Quando comparadas as taxas de mortalidade decorrentes da infecção, Santa Catarina apresentou uma considerável redução na taxa de mortalidade ao longo dos anos, passando de 9,1 em 2009 para 6,0 em 2018. Contudo o valor observado permanece acima da taxa de mortalidade observada no país (4,4 a cada 100 mil habitantes) para o mesmo período (Ministério da Saúde, 2019). Enquanto no estado catarinense observa-se que o número total de homens infectados (n=30.268) é bem superior ao número de mulheres

(n=19.312), representando 61% e 38,9%, respectivamente, no Brasil o número de mulheres infectadas até 2019 é inferior, representando apenas 34,4% dos casos notificados.

Outro fator contrastante entre a epidemia de Santa Catarina e a brasileira é a contribuição de categoria de exposição heterossexual para o crescimento da epidemia de HIV-1 em indivíduos do sexo masculino. Segundo o Ministério da Saúde (2019), em Santa Catarina a principal via de transmissão associada é a heterossexual (60,2% dos contágios), seguida pela categoria HSH (homens que fazem sexo com homens) (34,4%), enquanto no Brasil a população HSH e heterossexual contribuem quase que igualmente para o aumento do número de casos de HIV-1 notificados (39,2% vs. 39,1%). Ao longo dos anos contudo, foi possível observar que houve aumento na contribuição da categoria HSH, passando de 16,1% da forma de contágio relatada em 2007 para 27,4% em 2019 (Ministério da Saúde, Sistema de Informação de Agravos de Notificação, 2019).

1.3 Redes de transmissão viral

A diversidade de cepas virais circulantes dentro de cada indivíduo infectado é determinada por uma série de fatores intrínsecos ao vírus, tais como alta taxa de mutação, eventos de recombinação e elevada capacidade replicativa do vírus, além de questões envolvendo o hospedeiro como a pressão exercida pelo sistema imune e o uso de medicação antirretroviral (Wilbe, 2004). Devido a ação destas variáveis ao longo do período de infecção, a composição genética dos vírus circulantes é praticamente única para cada indivíduo, tornando-se mais diferenciada da cepa viral original que causou a infecção à medida que o tempo passa. Com base nessas características, a utilização de técnicas modernas de filogenia permite, a partir da amostragem e sequenciamento do genoma viral (completo ou parcial) de uma população, identificar quais indivíduos compartilham vírus geneticamente próximos. Essa similaridade é mais acentuada quando existe uma relação transmissor-receptor entre os indivíduos analisados, ou seja, é um indicativo que eles estejam ligados por um evento de transmissão. A partir desses dados é possível então reconstruir as chamadas redes de transmissão do HIV-1 (Hassan et al., 2017).

Atualmente, diversos estudos têm utilizado dados genéticos do HIV-1 para inferir as características dos indivíduos epidemiologicamente relacionados e assim entender os padrões de disseminação do patógeno em um determinado local. Em um estudo conduzido em Chicago (EUA), buscando elucidar o padrão de transmissão do HIV-1, evidenciou-se que a população HSH afrodescendente era mais susceptível à infecção, sendo necessária uma abordagem diferenciada de prevenção e acompanhamento para esse subgrupo (Lubelchek et al., 2015). No Reino Unido, o estudo das redes de transmissão revelou que a população HSH transmite HIV-1 de forma mais rápida que indivíduos heterossexuais, sendo que a transmissão ocorre de forma mais acentuada durante a fase aguda da infecção (Lewis et al., 2008). Mudanças nas dinâmicas de transmissão ao longo do tempo dentro da população HSH também foram observadas em estudos conduzidos na cidade de Detroit (EUA) (Volz e Frost, 2013) e na Holanda (Ratmann et al., 2016). Por sua vez, na África do Sul, a reconstrução das redes de transmissão evidenciou que a incidência exorbitante observada em jovens mulheres ocorre principalmente pelo contato sexual com homens de idade mais avançada (de Oliveira et al., 2016). Neste país, mulheres jovens são um grupo extremamente vulnerável ao HIV-1 e medidas de prevenção focadas nos indivíduos que transmitem HIV-1 para elas seriam muito mais eficazes.

No Brasil, apenas um estudo foi conduzido para caracterizar as redes de transmissão do HIV-1. Junqueira et al (2016) buscaram esclarecer as dinâmicas associadas exclusivamente à dispersão do HIV-1B, utilizando-se de sequências disponíveis em bancos de dados públicos, ou seja, dados disponibilizados por estudos realizados ao longo de vários anos e executados em lugares diversos. Essa estratégia gerou uma amostragem abrangente da epidemia brasileira, porém muito esparsa e com poucos dados clínicos e epidemiológicos disponíveis. Essa limitação no acesso de dados acabou minimizando a análise e resultados do estudo na identificação e interpretação, de um ponto de vista epidemiológico, das redes de transmissão no Brasil. Para que um estudo de redes de transmissão atinja seu potencial máximo em esclarecer os eventos e populações associadas à transmissão viral do HIV-1 torna-se vital que a amostragem seja densa, cobrindo diferentes

subpopulações expostas ao vírus (indivíduos heterossexual, população HSH e UDI, por exemplo), e com abundância de dados geográficos, clínicos e epidemiológicos, possibilitando identificar pares de indivíduos “receptores” e “doadores” do vírus presentes na amostragem. A partir da observação dos padrões de transmissão, pode-se então aperfeiçoar programas de vigilância, prevenção de novos casos e combate à doença.

Uma maneira de aumentar o volume de dados moleculares e epidemiológicos associados a epidemia de HIV-1, possibilitando a sua utilização na reconstrução das redes de transmissão, é a utilização de dados obtidos a partir de redes nacionais de genotipagem. Em países desenvolvidos, esse procedimento é realizado no momento do diagnóstico, previamente ao início do tratamento (Wang et al., 2019, Kantor et al., 2015). Essa medida, além de garantir a eficácia do tratamento, por indicar de antemão possíveis resistências, fornece também dados moleculares e epidemiológicos recentes pós-infecção, permitindo uma reconstrução das redes de transmissão ativas no momento, bem como identificar grupos em maior vulnerabilidade e envolvidos na disseminação viral. Em localidades onde essa medida não ocorre, como em países da África por exemplo, observa-se um intenso investimento em programas de pesquisa e monitoramento, almejando sequenciar uma parcela expressiva da população infectada, elucidando fatores e grupos associados aos crescentes números da epidemia (de Oliveira et al., 2016).

No Brasil, a Rede Nacional de Genotipagem (RENAGENO) procede a genotipagem de indivíduos infectados após a falha terapêutica. Embora a coleta e sequenciamento do genoma viral seja feito de forma posterior ao início do tratamento, e, conseqüentemente, após um longo período de diagnóstico e infecção, essa forma de abordagem ainda fornece dados suficientes para tentar reconstruir as redes de transmissão do país e identificar fatores capazes de melhorar as políticas de monitoramento e controle da epidemia. Além disso, após mudanças no protocolo de acesso ao tratamento e notificação de casos de HIV-1 no país, acredita-se que o número de sequências e pacientes amostrados pela RENAGENO irá aumentar substancialmente, aumentando a cobertura/amostragem populacional dos indivíduos infectados (Ministério da Saúde, 2017, 2018 e 2019).

1.4 Ferramentas para detecção de redes de transmissão

A identificação de redes de transmissão é um processo combinado de distintas etapas metodológicas. Essas redes nada mais são que sub-árvores/agrupamentos/*clusters* de sequências provenientes de indivíduos infectados e que compartilham um ancestral comum, ou seja, encontram-se relacionadas filogeneticamente (Dearlove et al., 2017; Hassan et al., 2017) (Figura 5). Os termos rede ou *cluster* de transmissão são utilizados indiscriminadamente.

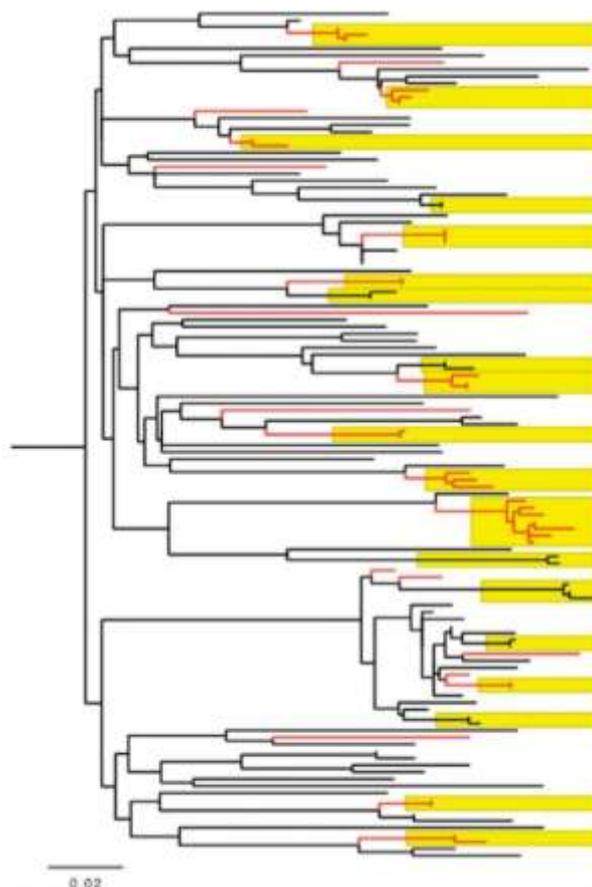


Figura 5: Representação esquemática de uma reconstrução filogenética e determinação de possíveis relações de transmissão. Adaptado de Dennis *et al.*, 2014.

Em amarelo encontram-se demarcados possíveis clados/agrupamentos de transmissão. Essas relações de transmissão aparecem representadas na forma de subárvores, onde os indivíduos que compõem esses clusters apresentam pequenas distâncias genéticas entre si (representada pelo tamanho curto dos ramos) e elevado suporte de nó (demonstrado pelas linhas em vermelho).

A primeira etapa do processo de identificação de redes de transmissão é estimar a relação evolutiva entre as sequências. Para isso torna-se necessário empregar um método filogenético de construção/inferência de árvores. Há duas formas de proceder essa reconstrução: utilizando métodos baseados em distância ou métodos baseados em caracteres e otimização. Os mais comumente utilizados são Neighbor-Joining (NJ), Máxima Verossimilhança (ML) ou Inferência Bayesiana (BI) (Grabowski e Redd, 2014; Hassan et al., 2017). O método de NJ é mais simples e consiste na utilização de um algoritmo onde as sequências alinhadas são agrupadas de acordo com a sua similaridade genética. É uma metodologia computacionalmente eficiente, passível de utilização para grande volume de dados, em vista que o processo de clusterização não envolve a busca todo o espaço amostral da árvore. Já os métodos de ML e BI, embora sejam computacionalmente menos eficientes, são abordagens mais complexas, com modelo probabilístico associado e capazes de resultar em árvores filogenéticas mais acuradas. Nessas duas metodologias, a investigação/busca do espaço amostral da árvore, para melhor reconstruir as interações filogenéticas, é mais extensiva. Além disso, torna-se necessário empregar conjuntamente modelos que descrevam a evolução viral como, por exemplo, modelos de substituição nucleotídica. A ML baseia-se na tentativa de maximizar a verossimilhança entre todas as árvores possíveis dado um conjunto de dados. O método probabilístico de BI procura buscar entre um conjunto de árvores mais prováveis a que melhor se adequa aos dados disponíveis, contudo, para isso, a busca é realizada em um espaço amostral limitado por informações determinadas previamente (*priors*). Quanto mais informações prévias adicionadas, maior a capacidade desse modelo em reconstruir com maior acurácia as relações filogenéticas (Grabowski e Redd, 2014; Hassan et al., 2017).

Com vista em identificar redes de transmissão, independente do método de reconstrução filogenética, é importante a associação de um cálculo de suporte de ramo. Os mais utilizados são o bootstrap, a probabilidade posterior (PP), o teste de razão de verossimilhança aproximada (aLRT) e, mais recentemente, o *transfer bootstrap expectation* (TBE) (Brenner et al., 2007; Lewis et al., 2008; Hué et al., 2009; Yerly et al., 2009, Hassan et al., 2017,

Grabowski e Redd, 2014, Lemoine et al, 2018). Essa etapa é crucial para a determinação de um *cluster* de transmissão, pois é a partir da avaliação de ponto de corte desse parâmetro que os *clusters* passam a serem selecionados. Após essa seleção prévia, a etapa seguinte consiste em determinar um limite de distância genética (Prosperi et al., 2011; Ragonnet-Cronin et al., 2013). É a partir desse segundo ponto de corte que serão selecionados as relações filogenéticas com possibilidade de representarem contatos epidemiológicos (Grabowski e Redd, 2014; Hassan et al., 2017). Embora todas as sequências agrupadas em *clusters* de transmissão apresentem o mesmo ancestral comum, nem todo agrupamento de sequências que compartilham dessa característica são considerados epidemiologicamente relacionados. Um exemplo disso é a relação filogenética observada para os subtipos B e C do HIV-1. Quando realizada uma inferência filogenética, todas as sequências classificadas como subtipo HIV-1C agrupam-se juntas, em um clado monofilético separado do subtipo B. Embora as sequências formem um *cluster* de alto suporte, é sabido que nem todos esses indivíduos amostrados encontram-se epidemiologicamente relacionados por eventos de transmissão viral.

A etapa de determinação de ponto de corte de distâncias genética é de tamanha importância que há ferramentas desenvolvidas para determinação de agrupamentos (ex. HIV-TRACE) (Pond et al., 2018) que utiliza-se exclusivamente desse recurso para inferir *clusters* de transmissão. Embora essa etapa seja de suma relevância para o processo de determinação das redes de transmissão, a escolha do ponto de corte normalmente é determinada pelo operador, podendo variar significativamente entre os trabalhos e afetando significativamente a seleção de *clusters* (Lemoine et al., 2018; Junqueira et al., 2019).

Embora inúmeras metodologias tenham sido desenvolvidas para identificar e caracterizar as redes e eventos de transmissão, ainda há controvérsias quanto às formas de amostragem, abordagens metodológicas e alvos genômicos utilizados nas reconstruções (Junqueira et al., 2019). A falta de padronização nessas análises provocou um aumento de interesse em estabelecer como a escolha do fragmento gênico e o seu tamanho influenciam

no agrupamento filogenético do HIV-1 (Harris et al., 2003; Novitsky et al., 2015).

1.4.1 Influência da determinação de ponto de corte para suportes de ramo e distância genética

Embora existam diferentes abordagens para determinar o suporte de ramo já mencionadas nesse trabalho (Brenner et al., 2007; Lewis et al., 2008; Hué et al., 2009; Yerly et al., 2009, Hassan et al., 2017, Grabowski e Redd, 2014, Lemoine et al., 2018), a técnica tradicionalmente utilizada é o *bootstrapping* (Efron et al., 1996). Durante o *bootstrapping*, os sítios do alinhamento (cada posição) são randomicamente selecionados e embaralhados formando novas pseudo-replicatas do alinhamento. A partir desses novos alinhamentos, são montadas árvores filogenéticas e observadas com que frequência essas pseudo-replicatas se organizam em *clusters*. A árvore final é obtida pelo consenso das análises. Devido a isso, o *bootstrap* pode ser considerado um método de reprodutibilidade. Há também outros métodos que procuram avaliar o agrupamento dos dados utilizando-se de ferramentas estatísticas com, por exemplo, a probabilidade posterior (PP) e o TBE (Lemoine et al., 2018), sendo considerados métodos mais robustos que o *bootstrap* (Alfaro et al., 2003, Erixon et al., 2003, Lemoine et al., 2018).

Há descrito intervalos de valores de parâmetros para definições de cluster em HIV-1 onde o suporte de ramo por *bootstrap*, SH-like e PP podem oscilar entre 50 a 100 (Hué et al., 2004; Kaye et al., 2008; Cuevas et al., 2009; Chalmet et al., 2010, Bezemer et al., 2010; Pilon et al., 2011; Marzel et al., 2016). Geralmente, a abordagem combinada entre o uso de um suporte de ramo e um ponto de corte de distância é a mais utilizada; entretanto alguns estudos já foram conduzidos utilizando-se apenas de valor de suporte, como revisado por Junqueira et al (2019). O alto suporte estatístico do ramo contudo pode não garantir que os membros desse clado sejam necessariamente relacionados epidemiologicamente entre si (Felsenstein et al., 1985). Devido a isso, a associação da utilização de pontos de corte no suporte de ramo e na distância genética auxilia na exclusão de possíveis *clusters* de transmissão

falsos, especialmente quando o período de amostragem ou a área de abrangência do estudo são grandes (Junqueira et al., 2019).

Outro parâmetro importante na detecção de clusters de transmissão é a distância genética. Este parâmetro consiste na proporção de sítios diferentes entre duas sequências, ou seja, é baseado no número de substituições nucleotídicas observadas entre duas referências. Pode-se também utilizar um modelo de substituição nucleotídica para corrigir a distância observada considerando mutações intermediárias ou reversões. Alternativamente, pode-se calcular a distância patrística para estimar-se a distância genética entre dois vírus. Este parâmetro é o somatório do comprimento dos ramos que ligam/conectam dois nodos (externos ou internos) em uma árvore filogenética, sendo que neste caso, o tamanho do ramo já é estimado com base no modelo evolutivo de escolha (Fourment e Gibbs, 2006).

Novamente, há uma grande variação de pontos de corte associados à distância genética entre sequências dentro de um mesmo cluster, variando de 1 a 9% (Kaye et al., 2008; Hughes et al., 2009; Chalmet et al., 2010; Prospero et al., 2011; Mehta et al., 2012; Junqueira et al., 2016; Yebra et al., 2018) ou 0,020 a 0,080 quando medida em distância patrística (Prospero et al., 2011; Bezemer et al., 2015; Poon et al. 2016). Entretanto, a não padronização desses valores ou a escolha deles pelo analista sem um racional lógico leva a uma grande variação na quantidade e tamanho dos *clusters* encontrados na população estudada, influenciando diretamente nos resultados do estudo (Figura 6).

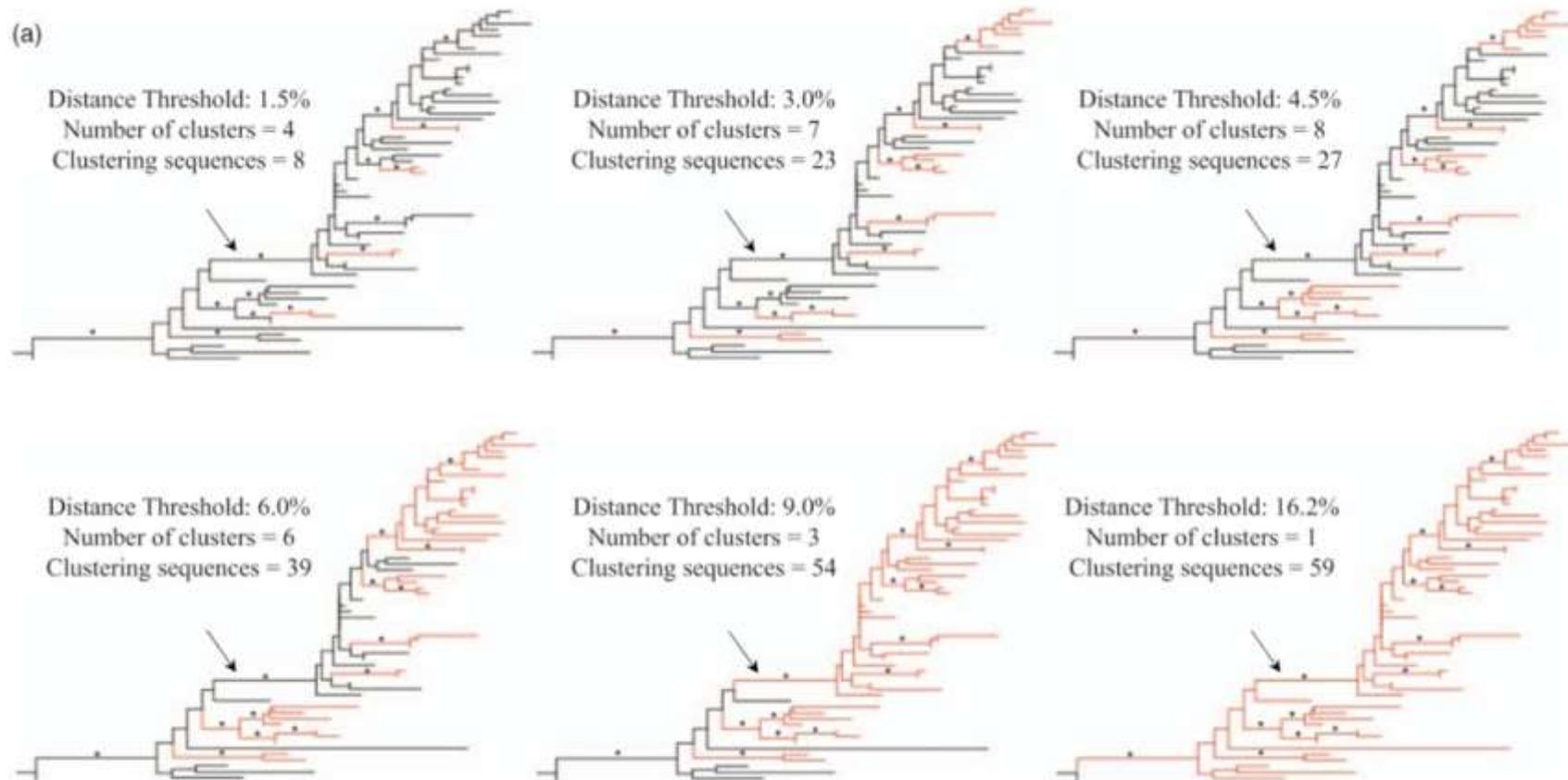


Figura 6: Representação da influência da determinação de ponto de corte de distância genética na quantidade e tamanho dos *clusters* de transmissão identificados. Adaptado de Hassan *et al.*, 2017.

A partir da reconstrução filogenética apresentada é possível observar que para análises restritivas, nas quais é utilizada uma distância genética pequena, o número de *clusters* identificados é grande, contudo o tamanho e número de indivíduos amostrados é pequeno. Ao fazer uma análise mais permissiva, com aumento progressivo do valor máximo de distância genética aceita entre os indivíduos, o número de *clusters* é diminuído, contudo há aumento no número de indivíduos amostrados e tamanho dos agrupamentos identificados.

Marzel et al (2016) conduziu estudo avaliando o efeito de 104 combinações de distâncias genéticas máximas intracluster (variando de 1 a 2,5%) e suporte por *bootstrap* (variando de 50 a 100% com 2% de incremento), observando que quanto maior a distância genética utilizada, maior o tamanho dos clados. Esse achado é sugestivo de que tanto a distância quanto o suporte de ramo possuem um grande impacto na identificação das relações de transmissão.

A distância genética também é importante na datação dos *clusters* de transmissão. O ponto de corte de distância utilizado na análise pode priorizar a identificação de redes mais recentes e ativas ou retratar uma rede antiga. Quando a análise é mais restrigente, ou seja, utiliza-se de ponto de corte de distância menores, aumenta-se a probabilidade de identificar casos relacionados a eventos recentes e rápidos de transmissão. Por exemplo, Wertheim e colaboradores (2017) identificaram que a distância genética de 2% era a máxima permitida para se detectar redes de transmissão de HIV-1 contendo apenas parceiros sexuais previamente identificados por entrevistas em Nova York (EUA). Além dessa distância, as redes começavam a misturar indivíduos sem histórico de contato sexual. Entretanto, a modulação dessa variável, aumentando seus valores, pode auxiliar na identificação de eventos de transmissão mais antigos (Division of HIV/AIDS Prevention & Centers for Disease Control, 2018). Por exemplo, para que todas as sequências de uma rede de transmissão viral bem caracterizada de usuários de drogas injetáveis na Finlândia permanecesse dentro do mesmo cluster, uma distância genética de 9% foi utilizada (Hassan et al., 2017). Esse fato deve-se ao longo período que se passou entre infecção e coleta para sequenciamento, permitindo com que a população viral de cada paciente divergisse bastante da cepa original.

Vários programas foram desenvolvidos para identificar *clusters* de transmissão. Um dos mais populares, devido a sua flexibilidade e capacidade de processar grandes *datasets*, é o Cluster Picker (Ragonnet-Cronin et al., 2013), que utiliza uma metodologia com abordagem combinada utilizando suporte de ramo e distância genética como critérios de seleção para os *clusters* em uma árvore filogenética. Seus arquivos de entrada consistem em uma

árvore filogenética previamente estimada e o alinhamento de sequências utilizadas para a reconstrução filogenética. O algoritmo inicia subdividindo a árvore baseada no ponto de corte do suporte de ramos determinado pelo usuário. Estas sub-árvores de alto suporte são então analisadas quanto a distância genética, par-a-par, entre as sequências que compõem aquele agrupamento. Se o valor de suporte e de distância genética de um grupamento de sequências está dentro dos valores determinado pelo usuário, o programa identifica aquele agrupamento como um cluster de transmissão.

Devido a dificuldade de se escolher os valores de ponto de corte, há cada vez mais estudos tentando utilizar abordagens que não exigem essa pré-definição de parâmetros. O GapProcedure (Vrbik et al., 2015) é uma abordagem automatizada de clusterização por distância que não necessita valores de ponto de corte determinados pelo operador. Essa metodologia está baseada em uma matriz de dissimilaridade construída utilizando-se de modelos populares de substituição nucleotídica. As sequências são comparadas entre si, par-a-par, e as lacunas nas distâncias ordenadas são utilizadas para sugerir quais indivíduos estariam epidemiologicamente relacionados. Entretanto, como o método não se baseia em árvores filogenéticas, os *clusters* identificados não necessariamente são monofiléticos, dificultando a interpretação dos mesmos como redes de transmissão.

Por sua vez, Barido-Sottani e colaboradores desenvolveram um método baseado no modelo *multi-state birth-death* (MSBD) que estima taxas de transmissão ao longo da árvore filogenética (Barido-Sottani et al., 2018). Esse modelo vem sendo amplamente utilizado para modelar a estrutura populacional e construir análises filogenéticas baseadas em indivíduos inseridos em uma estrutura de comunidade (Maddison et al., 2007; FitzJohn, 2012; Stadler e Bonhoeffer, 2013; Kühnert et al., 2016). Esse método assume que os *clusters* de transmissão são induzidos na comunidade através das redes de contato sexual. Quando o patógeno é inserido e transmitido em uma nova comunidade, a taxa de transmissão aumenta, em vista do elevado número de indivíduos não infectados e susceptíveis. O MSDB busca detectar variação nas taxas de transmissão de uma determinada reconstrução filogenética em escala de

tempo e os resultados dessa variação ao longo do tempo são interpretados como uma provável introdução de uma nova epidemia em uma população susceptível, indicada pela formação de um novo *cluster* (Barido-Sottani et al., 2018). Entretanto, a baixa processividade da atual implementação do método e a necessidade de se estimar árvores em escala de tempo dificultam muito a utilização desta metodologia em um grande conjunto de dados.

1.4.2 Influência da densidade amostral e uso para monitoramento de grandes epidemias e epidemias concentradas

Outro fator capaz de influenciar a detecção de redes de transmissão é a densidade amostral. Novitsky et al. (2014) promoveram a análise de sequências provenientes de duas fontes distintas, uma comunidade populacional da Botswana e sequências provenientes do Los Alamos HIV-1 Database, variando a porcentagem de sequências amostradas de 1 a 70%, nesses dois cenários. O agrupamento das sequências de HIV-1 revelou-se positivamente relacionada a densidade amostral. Ademais, amostragens abaixo de 10% foram associadas a um perfil de clusterização variável, indicando que resultados provenientes de uma baixa densidade amostral podem não refletir a realidade.

Estudos de grandes epidemias, abrangendo países e continentes, são usualmente caracterizadas por uma baixa densidade amostral dos indivíduos infectados e, conseqüentemente, as redes de transmissão nessas condições podem ser de difícil identificação. Para conseguir identificar os clusters de transmissão nesses casos torna-se importante flexibilizar as distâncias genéticas para abranger a maior diversidade genética encontrada na população, contudo o suporte de ramo associado aos clados deve permanecer alto. Além disso, em casos de baixa densidade amostral, é possível que muitos indivíduos intermediários no processo de transmissão (elos de ligação) estejam ausentes. Com isso, a distância genética entre os indivíduos amostrados torna-se maior, mas, mesmo assim, eles podem estar participando de uma rede de transmissão em comum (Poon, 2016).

1.4.3 Influência do tamanho e escolha de fragmento gênico

A escolha do fragmento e tamanho gênico a serem utilizados para as reconstruções filogenéticas e identificação dos *clusters* de transmissão ainda é bastante controversa (Hué et al., 2004; Jenwitheesuk e Liu, 2005; Stümer et al., 2005). A filogenia molecular baseada na utilização de apenas um gene pode apresentar resultados conflitantes e incongruências (Rokas et al., 2003; Rokas et al., 2008). Além disso, o uso de fragmentos pequenos podem ocasionar perda de sinal filogenético e surgimento de erros sistemáticos (Felsenstein et al., 1978; Jeffroy et al., 2006).

Outro fator importante associado à reconstrução é taxa evolutiva do gene escolhido. O gene da polimerase (*pol*) tem sido predominantemente utilizado em inúmeros estudos (Hué et al., 2004; Brenner et al., 2007; Volz et al., 2012; Volz et al., 2013; Wertheim et al., 2013; Junqueira et al., 2016); contudo o gene do envelope (*env*) também pode ser utilizado, em vista de apresentar um sinal filogenético mais robusto, ou seja, conseguem indicar relação de ancestralidade entre elas (DeBry et al., 1993; Zhang et al., 1993; Strunnikova et al., 1995).

Estudos prévios indicam que o grau de clusterização do HIV-1 está diretamente associado ao tamanho da sequência gênica utilizada para a reconstrução filogenética. Em estudos combinando diferentes métodos e tamanhos de fragmentos foi possível observar que comprimentos maiores de sequência identificam melhor os clados de transmissão, visto que disponibilizam uma quantidade maior de sítios informativos e comparativos para a análise (Lemey e Vandamme, 2005; Novitsky et al., 2015; Yebra et al., 2016).

Quanto à determinação de qual segmento gênico a ser utilizado, é possível observar que a maioria dos estudos elucidando redes de transmissão do HIV-1 utilizam-se de sequências do gene *pol*, principalmente devido a maior disponibilidade desse fragmento em função das análises de mutação de resistência associado ao uso de terapia antirretroviral (Junqueira et al., 2019). Novitsky et al. (2015), a partir do genoma total do HIV-1, inferiram que a

proporção de sequências de *pol* agrupadas foi maior que a de sequências de *gag* (em todos os bootstrap testados) e *env* (em bootstrap $\geq 90\%$). Além disso, quando foram comparados os agrupamentos utilizando-se subregiões genômicas (porção de um gene), a região que codifica a enzima transcriptase reversa (RT) apresentou a maior proporção de sequências agrupadas. Em consenso ao encontrado por Novitsky et al. (2015), Lemey e Vandamme (2005) encontraram que para fragmentos com tamanhos entre 800 pb e 1200 pb, houve aumento nos suportes de ramo e suportes gene-específico onde os segmentos 3' da *pol* conseguiram reconstruir bem os clados de transmissão. Os resultados desse estudo indicam uma associação positiva entre a proporção de clusterização e parâmetros como tamanho de fragmento, número de sítios variáveis e informativos, bem como a adequação do fragmento de *pol* para reproduzir a análise filogenética em comparação aos dados obtidos em genoma total.

Embora o estado de Santa Catarina esteja contido na região de maior prevalência de HIV-1 no país e ter Florianópolis como primeira capital nacional em número de novos casos da doença, pouco ainda é conhecido sobre a dinâmica de transmissão viral em seu território. Embora o estado invista em campanhas de prevenção e vigilância, o avanço da epidemia evidenciado pela disseminação para o interior do território e a detecção de novos casos evidencia que aplicação de novas abordagens são necessárias para melhor compreender quais subpopulações estão mais expostas e susceptíveis à infecção. Neste sentido, a investigação das redes de transmissão do HIV-1 através de dados genéticos do vírus é uma abordagem inovadora e que vem mostrando-se muito útil para revelar informações sobre a dinâmica da epidemia de aids. Nesse trabalho foi estudado um grande banco de dados composto de sequências biológicas e informações epidemiológicas de pacientes que passaram pelo serviço de genotipagem do HIV-1 em Santa Catarina. A população amostrada neste banco é formada por diferentes grupos de exposição e a partir da reconstrução das redes de transmissão do HIV-1 nesta amostragem espera-se caracterizar a dinâmica da epidemia no estado. Este tipo de informação pode ser usado para ampliar, melhorar e reformular

programas de saúde pública de forma mais direcionada, vislumbrando a redução de novos casos, diagnóstico e tratamento precoce desses indivíduos.

2. OBJETIVOS

OBJETIVO GERAL

Estudar a dinâmica de transmissão do HIV-1 em diferentes grupos populacionais e regiões geográficas no estado de Santa Catarina.

OBJETIVOS ESPECÍFICOS

- Identificar redes de transmissão do HIV-1 a partir de sequências virais de pacientes amostrados em Santa Catarina.
- Caracterizar epidemiologicamente os indivíduos que compõem as redes de transmissão.
- Avaliar a abrangência geográfica das redes de transmissão através de diferentes municípios de Santa Catarina.

3. O BANCO DE DADOS E ASPECTOS ÉTICOS

Os dados analisados nesse trabalho são oriundos da Rede Nacional de Genotipagem do HIV-1 (RENAGENO) e foram cedidos pela Departamento de Vigilância, Prevenção e Controle das IST, do HIV/aids e das Hepatites Virais do Ministério da Saúde. A RENAGENO é responsável por monitorar pacientes soropositivos em acompanhamento clínico que apresentaram falha terapêutica. Neste estudo foram analisadas todas as sequências de pacientes residentes em cidades do estado de Santa Catarina que foram genotipados entre os anos de 2008 e 2017 (RENAGENO-SC). Além das sequências nucleotídicas do HIV-1, foram também fornecidos dados epidemiológicos e clínicos dessa população, tais como: ano de coleta da amostra, localidade de residência do paciente, contagem de carga viral e células CD4+, sexo, escolaridade, data de nascimento e etnia. Os indivíduos da RENAGENO-SC amostrados nesse banco de dados são identificados por código alfanumérico, sendo impossível sua identificação e garantindo assim anonimato e sigilo absolutos. Esse trabalho foi submetido e aprovado pelo Comitê de Ética em Pesquisa do Instituto Gonçalo Moniz/Fiocruz-BA sob registro nº 15300719.5.0000.0040.

CAPÍTULO 1 – Análise temporo-espacial da diversidade genética e epidemiologia molecular do HIV-1 em pacientes em falha terapêutica do estado de Santa Catarina, região sul do Brasil.

4. METODOLOGIA

4.1 Controle de qualidade e subtipagem das sequências RENAGENO-SC

Foi procedido a análise de controle de qualidade nas sequências obtidas. Para isso, elas foram submetidas a ferramenta *Quality Control* do site *Los Alamos HIV Database* (disponível em: <http://www.hiv.lanl.gov/content/sequence/QC/index.html>). Essa ferramenta permite identificar sequências hipermutadas e indicar possíveis contaminações na reação de sequenciamento por identificar sequências idênticas. Sequências que não passaram no controle de qualidade foram excluídas do banco de dados e das análises posteriores. Além disso, apenas a sequência mais antiga daqueles pacientes com repetida amostragem foi utilizada.

Para proceder a subtipagem (Figura 7), todas as sequências foram submetidas ao programa REGA Subtyping Tool v3.0 (Peña et al., 2013) e ao COMET (Struck et al., 2014), sendo classificadas nos subtipos do grupo M do HIV-1. Para sequências que apresentaram discordância entre os subtipos determinados por essas duas ferramentas, a classificação foi confirmada através da construção de árvores filogenéticas. Para tanto, foram selecionadas sequências-referência para todos os subtipos e CRFs do HIV-1 disponíveis no banco de sequências *Los Alamos HIV Database* (<http://www.hiv.lanl.gov/>) e alinhadas às sequências da RENAGENO em estudo com o algoritmo *MAFFT* (Kato et al., 2002). A qualidade dos alinhamentos foi checada visualmente no programa Aliview (Larsson, 2014) e editada quando necessário, fazendo-se a remoção de gaps. Sequências com tamanho inferior a 75% da maior sequência do alinhamento foram excluídas. Árvores filogenéticas foram inferidas através do método de máxima verossimilhança no programa IQ-TREE (Nguyen et al., 2015), utilizando-se a plataforma CIPRES *Science Gateway* (Miller et al., 2010). As sequências foram então classificadas conforme seu agrupamento em alto suporte (ultra-fast bootstrap acima de 90) com sequências referência. As árvores foram visualizadas utilizando-se o programa Figtree v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Quando a inferência filogenética permaneceu inconclusiva, foi realizada

a avaliação padrão de recombinação das sequências através da análise de *bootscanning* no programa Simplot v3.5.1 (Lole et al., 1999). Para isso, janela de 250 pares de bases (pb) com passo de 30 pb foi utilizada. Também foi realizada a submissão no JPHMM quando o padrão de recombinação inferido pelo Simplot v3.5.1 foi inconclusivo.

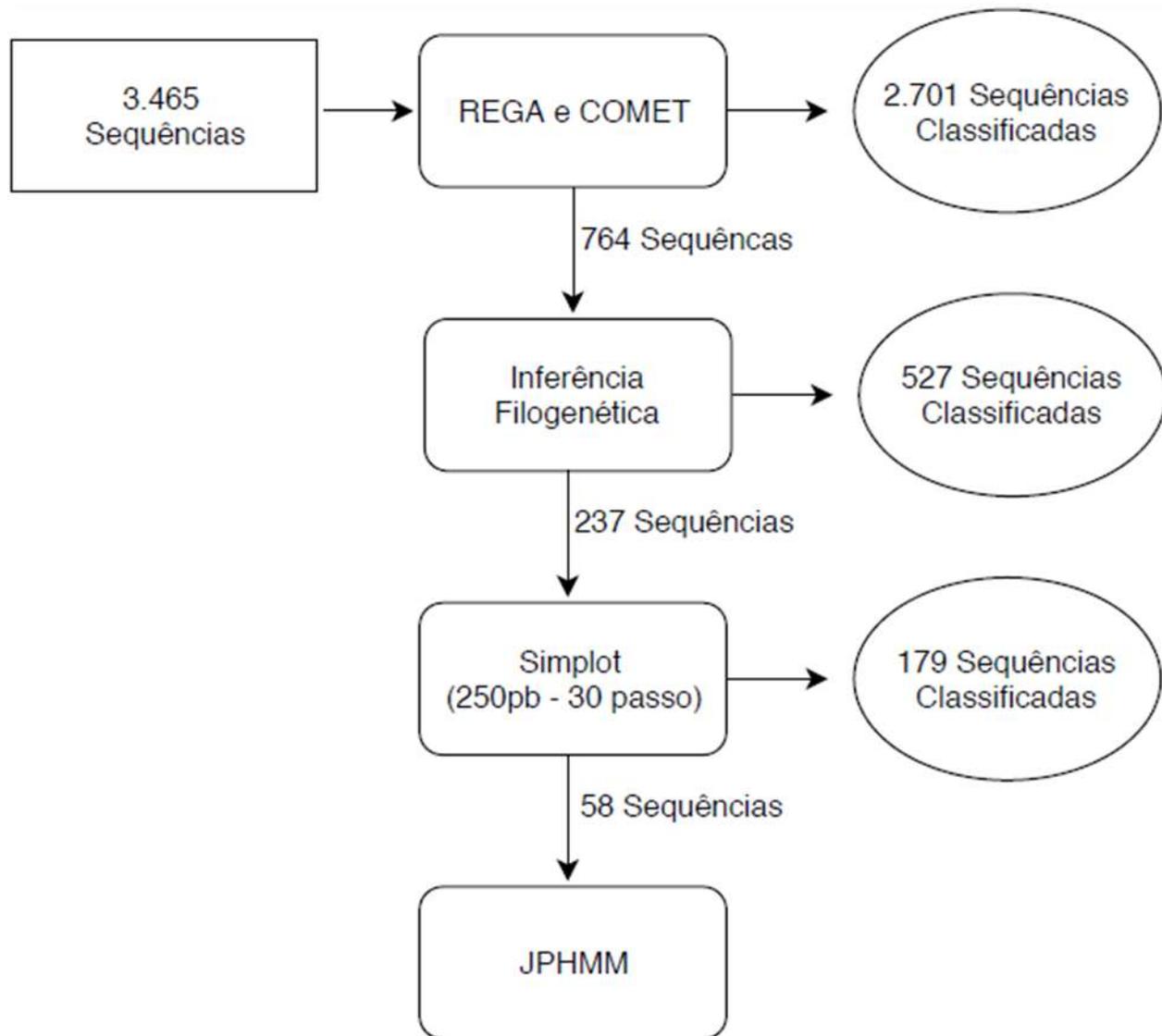


Figura 7: Fluxograma de metodologias aplicadas para subtipagem das sequências da RENAGENO-SC.

4.2 Análises Estatísticas

Para avaliar variáveis qualitativas tais como subtipo viral, gênero, etnia, ano de coleta, escolaridade e região intermediária de residência foram utilizados o teste de qui-quadrado ou Odds-Ratio. Para identificar tendências temporais como o aumento da prevalência dos subtipos durante o período de estudo foi utilizado o teste de regressão linear. As análises estatísticas foram executadas no programa R (disponível em: <http://R-project.org/>).

5. RESULTADOS

5.1 Caracterização molecular da epidemia de HIV-1 no estado de Santa Catarina

Após a subtipagem do gene da *pol*, das 3.070 sequências componentes do banco de dados RENAGENO-SC, 1998 (65,1%) foram classificadas como subtipo C, 597 (19,4%) como subtipo B e 149 (4,9%) como subtipo F1. O padrão de recombinação característico da CRF31_BC foi identificado em 112 (3,6%) amostras. Foram encontradas também 214 (7,0%) sequências classificadas como “Outras”. Nesse grupo foram incluídas formas do HIV-1 de menor relevância epidemiológica e que durante o período de coleta somaram separadamente menos de 1% de prevalência. Embora as formas recombinantes únicas entre os subtipos B e C do HIV-1 (URFs_BC) tenham sido descritas como responsáveis por 2,8% das infecções, foram também incluídas no grupo “outras” por não formarem um clado monofilético e possuírem origens diversas.

Durante o período amostragem da RENAGENO, foram encontradas também formas recombinantes complexas denominadas CPX, formadas por três subtipos distintos do HIV-1, tais como CPX_BCD (0,07%) e CPX_BCF1 (0,10%), além de outras URFs (formas recombinantes únicas) menos descritas, formadas pela recombinação do subtipo B e D (0,07%), B e G (0,03%), D e F1 (0,03%), recombinantes únicos envolvendo o subtipo C e F1 (0,13%) e uma maior prevalência de recombinantes únicos resultantes do subtipo B e F1 (0,85%). Quando analisadas as formas recombinantes circulantes com baixa representatividade foi observada uma maior prevalência de potenciais CRFs envolvendo os subtipos B e F1, sendo a CRF12_BF e a CRF71_BF1 responsáveis por 0,3% e 0,33% das infecções, respectivamente.

Ao longo dos dez anos analisados nesse estudo foi possível observar que o subtipo C está consistentemente aumentando em prevalência ($p < 0,01$, $r = 0,77$), crescendo de 56,9% em 2008 para 66,3% em 2017 (Figuras 8 e 9). Contudo, em direção oposta, as variantes raras classificadas no grupo “outras” apresentaram tendência à diminuição com significância estatística limítrofe

($p=0,0504$, $r= -0,63$), reduzindo de 7,9% para 5,4% em prevalência no mesmo período (Figuras 8 e 9). Para os demais subtipos nenhuma tendência temporal de mudança na frequência foi observada.

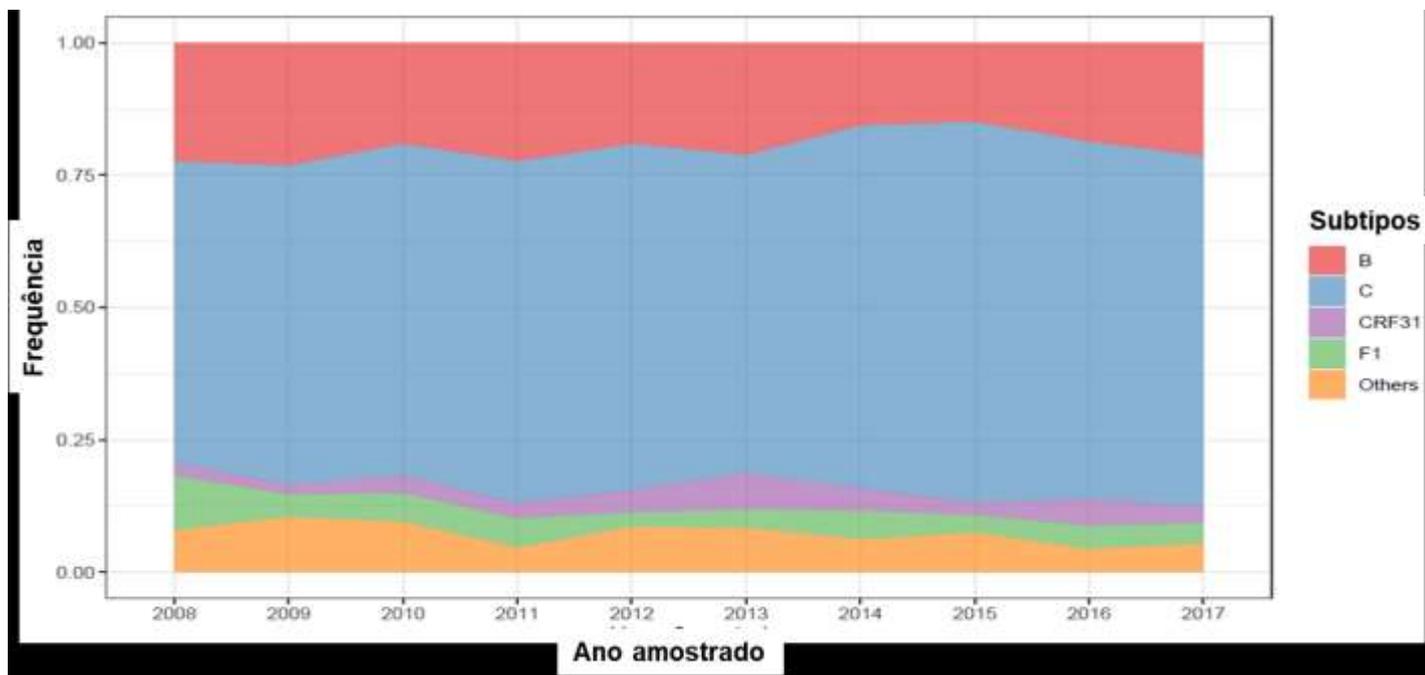


Figura 8: Variação anual da frequência dos subtipos de maior relevância epidemiológica descritos em pacientes com falha terapêutica no estado de Santa Catarina.

O subtipo C é o principal responsável pelas infecções relatadas em pacientes em falha terapêutica no estado de Santa Catarina, seguido pelo HIV-1B. Embora a frequência do subtipo mostre variações ao longo dos anos, é possível observar que o HIV-1C e as variantes raras (*Others*) apresentaram, respectivamente, um indício de aumento e diminuição. Durante o período, o subtipo C aumentou em prevalência de 56,9% para 66,3%, enquanto as variantes raras reduziram de 7,9% para 5,4%.

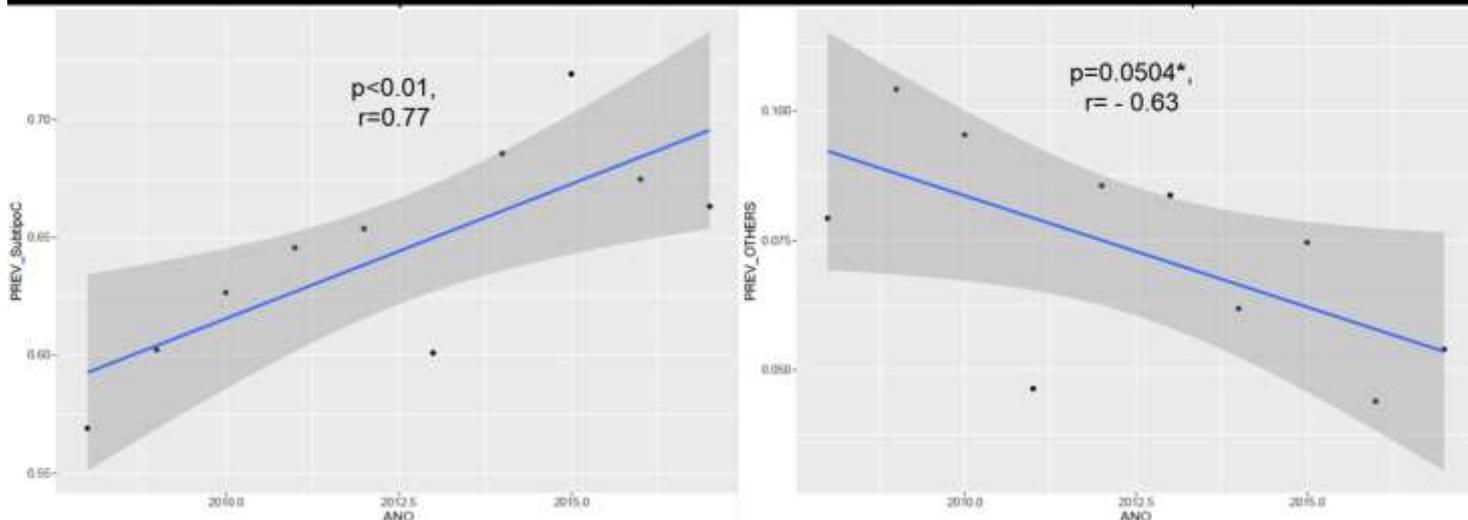


Figura 9: Regressão linear de frequência do subtipo no ano (eixo y) versus data de coleta (eixo x).

O indício de aumento observado para o subtipo C foi confirmado pelo teste de regressão linear e mostrou-se significativo ao longo do tempo ($p < 0,01$, $r = 0,77$), enquanto o decréscimo observado para outros subtipos (variantes raras) apresentou uma significância limítrofe ($p = 0,0504$, $r = -0,63$). Esse resultado é um indicativo de que, embora haja variação, o aumento e redução são constantes ao longo do período de estudo.

5.2 Características epidemiológicas dos pacientes da RENAGENO-SC

Das 3.070 sequências incluídas neste estudo, 51,1% foram provenientes de indivíduos do sexo masculino. A população amostrada da RENAGENO-SC é majoritariamente branca (85,3%), com idade entre 41 a 59 anos (40,7%) e com baixo grau de escolaridade, onde 44,4% dos indivíduos relataram possuir de 4 a 7 anos escolares completos (Tabela 1).

Ao relacionar a prevalência de subtipos encontrados e os dados epidemiológicos associados a essas sequências (Tabela 1), foi possível observar que 61,2% das infecções pelo HIV-1B ocorre em homens, enquanto o subtipo C mostrou-se relacionado à infecção em indivíduos do sexo feminino (52,1%) ($p < 0.001$). Ademais, o subtipo C e a forma CRF31_BC foram comumente observados em indivíduos com baixo nível educacional (4 a 7 anos

de estudo) quando comparado ao subtipo B, associado a indivíduos com nível educacional elevado (>12 anos) ($p < 0,001$). Esses dados também indicaram que o HIV-1B foi mais frequente em indivíduos com idades entre 25 e 40 anos, enquanto o subtipo C circula mais intensamente na população jovem (idade ≤ 24 anos) ($p < 0,001$). Nenhuma associação estatística relevante entre etnia e os diferentes subtipos do HIV-1 avaliados foi encontrada ($p = 0,38$).

Tabela 1: Dados demográficos referentes as quatro principais formas de HIV-1 encontradas no banco de dados RENAGENO-SC.

	HIV-1 Subtipos				Total	P value
	B	C	F1	CRF31		
Gênero						
Feminino	232 (38,9)	1041(52,1)	67 (45)	56 (50)	1396 (48,9)	< 0,001
Masculino	365 (61,2)	957 (47,9)	82 (55)	56 (50)	1460 (51,1)	
Etnia						
Negro	35 (6,6)	121 (7,1)	4 (3,0)	5 (5,2)	165 (6,7)	0,38
Caucasiano	451 (84,9)	1435 (84,8)	121 (92,4)	83 (86,5)	2090 (85,3)	
Outras etnias	45 (8,5)	137 (8,1)	6 (4,6)	8 (8,3)	196 (8,0)	
Anos de Escolaridade						
Nenhum	12 (2,6)	55 (3,7)	3 (2,5)	3 (3,6)	73 (3,4)	< 0,001
1-3 anos	48 (10,2)	190 (12,9)	16 (13,4)	7 (8,4)	261 (12,2)	
4-7 anos	171 (36,5)	689 (46,7)	48 (40,0)	46 (55,5)	954 (44,4)	
8-11 anos	149 (31,9)	429 (29,1)	44 (36,6)	20 (24,1)	642 (30,0)	
12 ou mais anos	88 (18,8)	112 (7,6)	9 (7,5)	7 (8,4)	216 (10,0)	
Idade						
<14 anos	42 (7,1)	221 (11,1)	10 (6,7)	10 (8,9)	283 (9,9)	< 0,001
15-24 anos	29 (4,8)	168 (8,4)	10 (6,7)	10 (8,9)	217 (7,6)	
25-40 anos	261 (43,7)	713 (35,7)	60 (40,3)	45 (40,2)	1079 (37,8)	
41-59 anos	251 (42,1)	806 (40,3)	64 (42,9)	41 (36,6)	1162 (40,7)	
>59 anos	14 (2,3)	90 (4,5)	5 (3,4)	6 (5,4)	115 (4,0)	

*Números e frequências em negrito destacam onde número de casos é significativamente maior ou menor que o esperado pela hipótese nula.

5.3 Dispersão heterogênea de subtipos de HIV-1 nas regiões intermediária de Santa Catarina

A partir do ano de 2017, o Instituto Brasileiro de Geografia e Estatística (IBGE) promoveu alterações na divisão regional do Brasil em regiões geográficas imediatas e intermediárias (IBGE, 2017). O estado de Santa Catarina passou a ter seus 295 municípios divididos ou alocados em 24 regiões imediatas e 7 regiões intermediárias, sendo elas Florianópolis, Caçador, Chapecó, Criciúma, Blumenau, Joinville e Lages. Para realizar a avaliação da dispersão de subtipos no estado catarinense, foi utilizada a divisão em regiões intermediárias.

Dos 295 municípios catarinenses, o presente estudo alcançou cobertura de 57,3%, amostrando pacientes em falha terapêutica de 169 municípios. As regiões intermediárias com maior cobertura foram Joinville, Criciúma, Florianópolis e Blumenau, com 80,0%, 79,5%, 76,5% e 75,0% dos municípios amostrados, respectivamente. As regiões de Lages, Caçador e Chapecó apresentaram as menores taxas de cobertura onde apenas 54,2%, 43,75% e 33% dos seus municípios foram amostrados, respectivamente.

O subtipo C mostrou-se o principal causador de infecções por HIV-1 em todas as regiões do estado, com prevalências de aproximadamente 70% nas regiões de Lages, Criciúma e Blumenau e prevalências próximas aos 60% nas regiões restantes (Figura 8). Já o HIV-1B foi o segundo em relevância epidemiológica para todas as regiões, sendo responsável por 19,4% de todas as infecções relatadas. O subtipo F1 mostrou-se causando infecção em 4,9% dos indivíduos avaliados. A forma recombinante CRF31_BC apresentou sua menor prevalência na região de Joinville (1,4%) e maior representatividade em Criciúma (5,9%), enquanto as variantes raras foram mais descritas na região de Caçador (13%), Joinville (9,5%) e Florianópolis (7,1%).

Apesar de uma homogeneidade geral entre as regiões de Santa Catarina quanto à frequência dos subtipos, o teste de qui-quadrado revelou uma leve variação territorial (Figura 10). O HIV-1B foi encontrado mais frequentemente na região de Joinville ($p < 0,001$), totalizando 28,1% das

infecções, enquanto o subtipo C e CRF31_BC foram os principais responsáveis por infecções nas regiões de Blumenau e Criciúma, correspondendo a aproximadamente 70% das infecções para o subtipo C e 5,9% para a forma recombinante em Criciúma. O subtipo F1 apresentou-se com elevada importância epidemiológica na região de Chapecó, onde foram relatados 12,4% de prevalência ($p < 0,001$). As regiões de Florianópolis e Caçador não apresentaram associação estatisticamente significativa para nenhum dos subtipos avaliados.

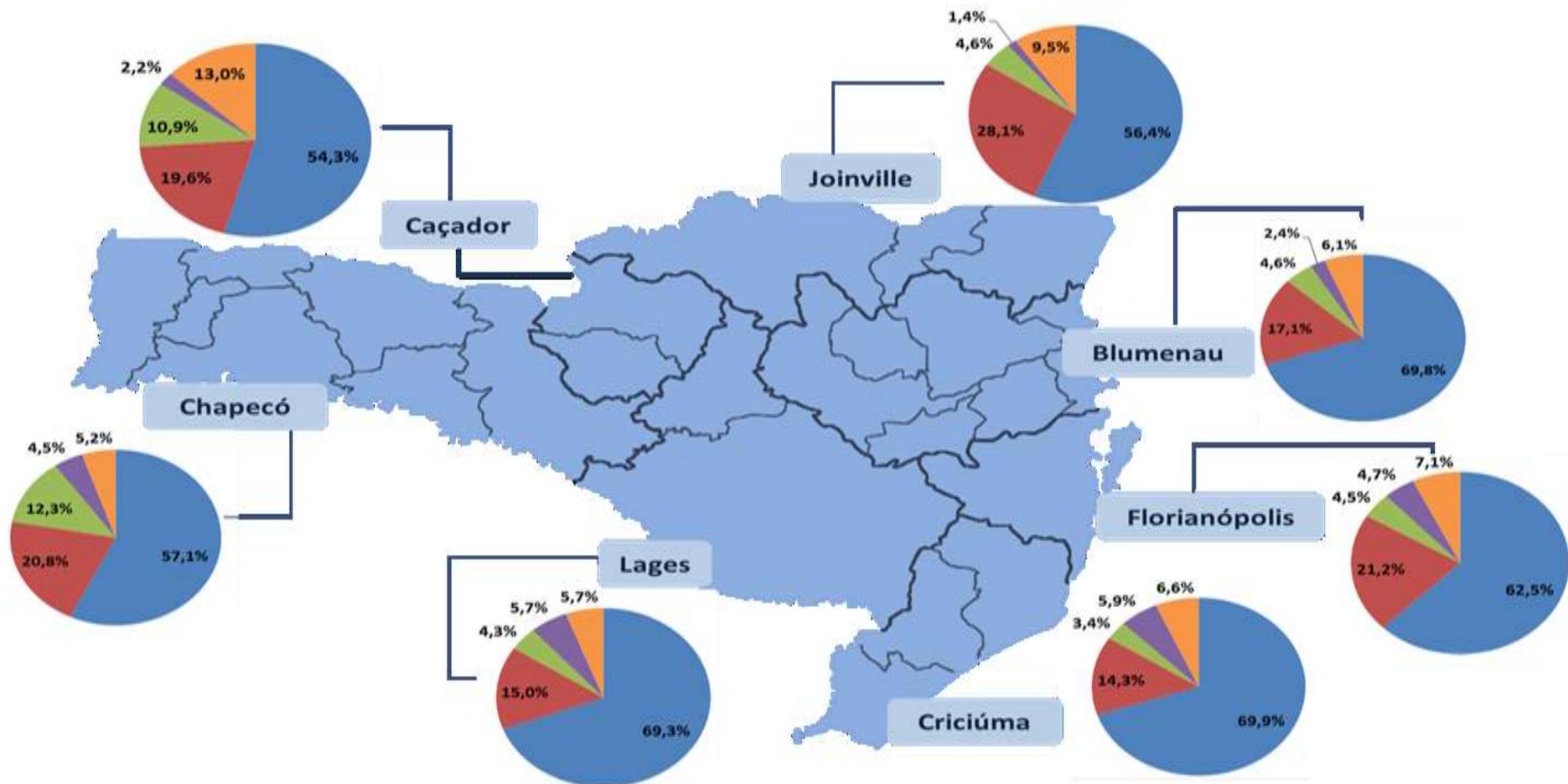


Figura 10: Diferenças na segregação espacial do HIV-1 B, C, F1 e CRF31_BC nas regiões intermediárias do estado de Santa Catarina.

Subtypes

- C
- B
- F1
- CRF31
- Others

O HIV-1C mostrou-se como maior responsável pelas infecções em todas as regiões observadas. Entretanto sua maior prevalência foi encontrada nas regiões intermediárias de Blumenau e Criciúma, atingindo aproximadamente 70% dos casos ($p < 0,001$). Os subtipos B e F1 apresentaram maior relevância epidemiológica nas regiões de Joinville, sendo responsável por 28,1% ($p < 0,001$), e de Chapecó, correspondendo a 12,3% ($p < 0,001$) das infecções, respectivamente. Não foram encontradas associações estatísticas significativas envolvendo a prevalência de subtipos e as regiões intermediárias de Florianópolis, Lages e Caçador.

6. DISCUSSÃO

Embora diversos estudos tenham sido conduzidos no estado de Santa Catarina procurando caracterizar molecularmente a epidemia da região e elucidar as características que a distinguem daquela encontrada no restante do país, a maioria foi conduzido nas cidades mais impactadas pelo HIV-1, não representando a população afetada como um todo. Este estudo pode comprovar a elevada prevalência do subtipo C em Santa Catarina, sendo o principal subtipo responsável pelas infecções no estado, seguido pelo HIV-1B. O presente trabalho corrobora resultados prévios, nos quais a prevalência relatada do subtipo C varia de 49 a 79% em amostras coletadas entre os anos de 2003 a 2014 (Schuelter-Trevisol et al., 2007; Locateli et al., 2007; Brígido et al., 2007; Rodrigues et al., 2010; Gräf et al., 2011; Gräf et al., 2016). Em estudos recentes conduzidos por Gräf e colaboradores (2016), o subtipo C revelou-se como responsável por 66% das infecções, enquanto o subtipo B contabilizou 20% das infecções. Foram relatadas também 8% de formas recombinantes únicas entre os subtipos B e C quando avaliado os genes *pol* e *env*, resultado da co-circulação de dois subtipos na mesma população (Gräf et al., 2016). Embora esse número seja bem expressivo, o presente estudo encontrou uma prevalência menor (2,8%) de URFs_BC, podendo essa diferença ser explicada pela análise exclusiva do gene *pol* no presente estudo e dessa forma não detectado recombinantes inter-gênicos.

Quando comparado aos outros estados do Sul, Santa Catarina apresenta a maior prevalência de subtipo C da região. No Rio Grande do Sul, o HIV-1C é responsável, em média, por aproximadamente 44% das infecções (Soares et al., 2003, Brindeiro et al., 2003, Soares et al., 2003b, Soares et al., 2005, Rodrigues et al., 2006, Santos et al., 2006, Monteiro et al., 2007, Brígido et al., 2007, Dias et al., 2009, Simon et al., 2010, Medeiros et al., 2011, Silveira et al., 2012, Gräf et al., 2016), enquanto no estado do Paraná essa prevalência varia de 21 a 30% (Brindeiro et al., 2003, Ferreira et al., 2008, Toledo et al., 2010, Raboni et al., 2010, Silva et al., 2010). Além disso, quando a relevância epidemiológica da forma recombinante CRF31_BC foi avaliada observou-se que apenas 3,6% dos casos de infecção por HIV-1 eram associadas a essa

forma viral, enquanto no Rio Grande do Sul, a CRF31_BC é responsável por 4% (Soares et al., 2005) a 35% (Simon et al., 2010) das notificações.

A elevada prevalência do HIV-1C na região sul e, principalmente, no estado de Santa Catarina, pode estar relacionado a elevada taxa de detecção e novos casos nessa região, sendo Florianópolis a primeira capital brasileira em taxa de detecção de aids (Ministério da Saúde, 2019). Alguns estudos sugerem que o subtipo C possui uma característica fenotípica menos virulenta em comparação a outros subtipos, mantendo a mesma capacidade de transmissibilidade (Abraha et al., 2009, Naidoo et al., 2019). O subtipo C utiliza-se predominantemente do co-receptor CCR5 para a entrada na célula, raramente trocando o tropismo para o co-receptor CXCR4. Já foi relatado que a mudança de tropismo para CXCR4 ou uso dos dois co-receptores está associado a uma progressão acelerada da doença (Vasan et al., 2006). Além disso, a redução do *fitness* replicativo e manutenção de níveis elevados de células CD4+ proporcionam que a progressão para o estágio de aids ocorra de forma mais lenta e gradual, fazendo com que os indivíduos infectados permaneçam assintomáticos por um longo período, com mais oportunidades de transmissão viral (Ariën et al., 2007, Abraha et al., 2009, Iordanskiy et al., 2010). Embora estudos relatem distinção entre a virulência do HIV-1C em comparação as outras variantes virais, Nunes e colaboradores (2014) em um estudo de coorte não encontraram diferenças significativas em relação a progressão clínica nos indivíduos infectados por diferentes subtipos. Entretanto, devido ao grande aumento da disponibilidade de antirretrovirais e aos novos protocolos que preconizam o início imediato do tratamento, estudos que acompanham a progressão natural da doença causada pelo HIV são cada vez mais raros, tornando muito difícil a elucidação de diferenças na patogênese causada pelos diferentes subtipos do HIV-1.

Ao avaliar retrospectivamente estudos conduzidos no estado de Santa Catarina, observa-se um aumento da prevalência do subtipo C, passando de 49% em amostras coletadas em 2004 (Locateli et al., 2007) para 66% em 2012-2014 (Gräf et al., 2016). Devido a amostragem ao longo de 10 anos, o presente estudo permitiu avaliar a variação temporal da prevalência dos

subtipos e, em consenso com outros relatos da literatura, comprovou que entre 2008 e 2017 o HIV-1C expandiu de 56,9% para 66,3% ($p < 0,01$, $r = 0,77$). Uma das explicações para o aumento da prevalência do subtipo C ao longo desse período poderia ser de fato um sucesso maior na sua transmissibilidade comparado aos outros subtipos (como discutido no parágrafo anterior). Alternativamente, pode ser devido a sua expansão para redes de transmissão distintas, como na população HSH. Essa hipótese é sugerida por Gräf et al. (2016), que por análises de filodinâmica observou um aumento da população efetiva do HIV-1 mesmo após a estabilização dos novos casos de infecção entre heterossexuais nos anos 2000 no sul do Brasil.

A partir do estudo apresentado aqui foi possível inferir que a razão de sexo (proporção de homens por mulheres infectados – M:F) nessa população é de 1,04, sendo inferior à relatada pelo Ministério da Saúde (2019) para o estado (1,9) e para o país (2,3). A razão de 1,04 indica que para cada homem infectado na população há também uma mulher infectada. A feminilização da epidemia já foi relatada no estado do Rio Grande do Sul e em outras unidades federativas do país, estando associada ao aumento da transmissão por via heterossexual (Brito et al., 2001, Fonseca et al., 2007, Simon et al., 2010). Além disso, o aumento do número de infecções em indivíduos do sexo feminino também pode estar associado a situações de vulnerabilidade social e de gênero, onde há recusa do parceiro no uso de preservativo e instabilidade financeira predispondo essas mulheres a infecção (Lazarini et al., 2012, de Matos et al., 2013). A evolução da razão de sexo observada para o estado segue o relatado para o país e similar ao estado do Paraná, onde houve redução da taxa de detecção de aids em mulheres de 2008 a 2018 (Ministério da Saúde, 2019).

Entretanto, o fato do banco de dados da RENAGENO-SC possuir razão de sexo perto de 1 é um forte indicativo de viés de amostragem nessa população, mostrando-se enriquecida de mulheres. A maior presença de indivíduos do sexo feminino pode ser decorrente de dois fatores: a execução de exames pré-natais e genotipagem de todas as gestantes infectadas com HIV-1, aumentando a detecção e disponibilidade de dados moleculares dessas

mulheres, além da a maior preocupação com a saúde demonstrada por essa população, caracterizada pelo acompanhamento de tratamento e detecção precoce de falha terapêutica. Devido às limitações decorrentes desse viés amostral, os resultados obtidos nesse trabalho devem ser interpretados com cautela. Além disso, a população de indivíduos amostrados pela RENAGENO é diferenciada quando comparada a outros estudos envolvendo infecção pelo HIV-1 conduzidos no estado e no país, limitando as comparações que podem ser realizadas baseadas em dados epidemiológicos gerais.

No presente estudo, apenas 6,7% dos indivíduos infectados são afrodescendentes e 8% de outras etnias (asiáticos e ameríndios). Os números descritos corroboram dados do IBGE sobre a composição étnica do estado de Santa Catarina, onde aproximadamente 83,9% dos indivíduos residindo no estado declararam-se brancos, 15,5% pardos e negros e 0,25% indígenas. (IBGE, 2010).

Ao relacionar a prevalência de subtipos encontrados e os dados epidemiológicos associados aos indivíduos, foi possível observar que HIV-1B e C estão associados a populações distintas. Infecções pelo subtipo B foram majoritariamente (61,2%) relatadas em homens, enquanto o HIV-1C foi mais presente em mulheres (52,1%) ($p < 0.001$). Embora não haja um consenso na literatura, diversos estudos epidemiológicos relatam associação entre infecção por subtipo C e indivíduos heterossexuais, enquanto o HIV-1B é encontrado com maior frequência em HSH (Brindeiro et al., 2003, Soares et al., 2005, Gräf et al., 2011, Raboni et al., 2010, Silva et al., 2010, Toledo et al., 2010, Silveira et al., 2012). O banco de dados de RENAGENO não fornece informações quanto a categoria de exposição, contudo as mulheres são majoritariamente infectadas pela via de transmissão heterossexual.

Houve, ao longo do tempo, um aumento no número de indivíduos infectados por via de exposição heterossexual, principalmente devido à feminilização da epidemia. Em 1990, no Brasil, a razão de sexo era de 5,4, decaindo para 2,3 em 2018 (Ministério da Saúde, 2019). Uma das hipóteses para rápida expansão do HIV-1C no sul do Brasil foi sua eficiente introdução e

disseminação no grupo heterossexual, permitindo que atingisse um elevado número de indivíduos e circulasse em grandes redes de transmissão. A segregação dos subtipos em subpopulações distintas deve-se muito provavelmente pela separação parcial das redes de transmissão onde esses subtipos foram introduzidos e se disseminaram nos primeiros momentos da epidemia. Entretanto, em estudo conduzido por Almeida et al. (2012) no qual foi avaliada a dinâmica temporal dos subtipos do HIV-1 em diferentes categorias de exposição na cidade de Porto Alegre (RS), observou-se que houve um aumento do subtipo C tanto na população HET quanto HSH entre os anos de 1998 e 2008, resultando em uma perda progressiva da associação entre subtipo e categoria de exposição. Isto mostra que mesmo redes de transmissão de diferentes subpopulações ou grupos de exposição, possuem interconexões e que a perda da associação subtipo-exposição é apenas uma questão de tempo, como já observada em outros países (Tovanabutra et al., 2004).

Contudo, Dias et al. (2009) descreveram uma possível diferença entre padrões de transmissão viral, onde haveria associação entre penetração anal e transmissão do HIV-B, não estando relacionado exclusivamente à exposição HSH. Esses achados sugerem que embora as redes de transmissão possam justificar a segregação, o subtipo C pode ser transmitido através de lesões menores, fator que pode ter contribuído positivamente para o sucesso na transmissão heterossexual (Dias et al., 2009).

Neste estudo os indivíduos infectados pelo subtipo B apresentavam idade entre 25 e 40 anos e alto nível de escolaridade (12 anos ou mais) ($p < 0,001$ para ambos). Em contraste, indivíduos acometidos pelo HIV-1C eram mais jovens (com idade inferior a 24 anos) e baixo nível escolar (4 a 7 anos de estudo) ($p < 0,001$ para ambos). A baixa escolaridade associada ao subtipo C pode estar relacionada a presença desse subtipo em indivíduos mais jovens (15-24 anos) e crianças abaixo de 14 anos ($p < 0,001$), correspondendo a aproximadamente 19,5% da população infectada por esse subtipo. A baixa idade apresentada por esses indivíduos faz com que apresentem um menor nível de instrução. Em contrapartida, a maior parte da população infectada com

subtipo B apresenta idade entre 25 a 40 anos (43,72%), proporcionando maior tempo para acesso educacional. Contudo, mais estudos são necessários para elucidar os fatores associados a essa diferenciação no acesso educacional desses indivíduos.

Quando avaliadas as regiões intermediárias instituídas pelo IBGE, foi possível observar que a cobertura de detecção de HIV-1 obtido no presente estudo foi menor que o descrito previamente por Dorsi et al (2004) (57,3% *versus* 71,3%, respectivamente). Essa diferença observada na baixa incidência de HIV-1 nos municípios catarinenses pode estar associada a uma subamostragem na região de Chapecó (33,0% das cidades amostradas), bem como na região de Caçador (43,75%) e Lages (54,2%). Além da variação de prevalência ao longo do tempo, este estudo revelou uma heterogeneidade na distribuição dos subtipos ao longo do território de Santa Catarina. A região de Joinville apresentou prevalência aumentada de subtipo B (28,1% das infecções), enquanto o HIV-1C e CRF31_BC foram predominantes nas regiões de Blumenau e Criciúma. O subtipo F1 mostrou-se relevante epidemiologicamente na região de Chapecó, correspondendo a 12,35% dos casos, sendo essa prevalência possivelmente associada à fronteira com a Argentina, onde essa forma viral tem sido amplamente relatada (Pando et al., 2011). Essas diferenças contrastantes na distribuição espacial podem estar associadas a influências e interconexões de epidemias de outras localidades, como, por exemplo, o intenso fluxo migratório entre as populações dos estados da região Sul, bem como a epidemia de fronteira. Joinville encontra-se ao norte do estado, próximo a fronteira com o Paraná, onde há baixa prevalência de subtipo C e predomínio de B (Brindeiro et al., 2003, Ferreira et al., 2008, Toledo et al., 2010, Raboni et al., 2010, Silva et al., 2010). Estudo conduzido por Raboni et al. (2010) indicou maior proporção de subtipo C (31%) em cidades localizadas próximas à fronteira com Santa Catarina, enquanto a prevalência relatada ao norte do estado do Paraná, próximo a divisa com o estado de São Paulo foi aproximadamente três vezes menor (13%). Tais dados corroboram a hipótese de que há uma alta interconexão entre as redes de transmissão de Santa Catarina e estados vizinhos, justificando possivelmente a elevada prevalência de HIV-1B encontrada na região de Joinville. Ademais, as regiões

de Criciúma e Lages, que apresentam aproximadamente 70% de infecções causadas por HIV-1C, encontram-se próximas a fronteira com o Rio Grande do Sul, cuja capital é considerada o epicentro de disseminação deste subtipo para outras regiões do Brasil (Gräf et al., 2016).

CAPÍTULO 2 – Análise da dinâmica de transmissão do HIV-1 em pacientes em falha terapêutica do estado de Santa Catarina, região sul do Brasil.

7. METODOLOGIA

7.1 Remoção de sítios associados a mutações de resistência aos ARV

Para a remoção dos principais códons associados à resistência, foram alinhadas no *MAFFT* (Kato et al., 2002) juntamente com as sequências da RENAGENO, sequências-guias completas dos genes da protease e da transcriptase reversa. Foram removidos os códons das posições 69, 72, 90, 96, 138, 141, 144, 150, 159, 162, 219, 228, 246, 249, 252, 255, 264, 270 da protease e 420, 492, 498, 504, 507, 519, 522, 528, 597, 600, 606, 615, 642, 645, 750, 834, 840, 849, 861, 867, 927, 942, 954, 972, 987 da transcriptase reversa. A escolha de remoção destes sítios específicos foi baseada em estudo conduzido por Bennett et al. (2009), onde foram determinados os principais sítios associados às mutações de resistência à terapia antirretroviral, a partir de estudo de vigilância epidemiológica. Esses sítios foram removidos previamente às análises de redes de transmissão para evitar viés de convergência evolutiva, ou seja, evitar que indivíduos sejam classificados como relacionados epidemiologicamente devido a pressão seletiva causada por histórico de tratamento semelhante.

7.2 Seleção de sequências externas brasileiras e internacionais

A fim de verificar se as redes de transmissão observadas em Santa Catarina se expandem para outros países e estados brasileiros, foram selecionadas sequências externas ao banco de dados RENAGENO-SC (Figura 11). Para seleção de sequências brasileiras foram obtidas todas as sequências cujo país de origem do paciente era o Brasil disponíveis no *Los Alamos HIV Database* (<https://www.hiv.lanl.gov/>). Foram baixadas sequências dos subtipos B, C e F1 (que foram as variantes mais frequentes entre sequências da RENAGENO-SC), com fragmento de tamanho entre 2262 e 3254, selecionando o segmento gênico da *pol*. Foi selecionada apenas uma sequência por paciente e excluídas sequências problemáticas conforme classificação do site. Sequências de Santa Catarina originadas de estudos passados não foram selecionadas devido ao fato de que o mesmo indivíduo poderia estar também incluído no banco de dados RENAGENO-SC.

Para seleção de sequências externas internacionais foram analisados os dados provenientes do Anuário Estatístico de Turismo de Santa Catarina (Secretaria Estadual de Turismo, Anuário Estatístico de Turismo de Santa Catarina, 2003 a 2017) para identificar os principais países de origem dos turistas que visitam o estado de Santa Catarina. Foram tabulados os dados compreendendo o período de 2003 a 2017 e selecionados países que, dentro desse período, contabilizaram mais de 1000 entradas no estado. Foram selecionados os seguintes países: Alemanha, Argentina, Austrália, Bolívia, Canadá, Chile, Colômbia, Espanha, Estados Unidos da América, França, Inglaterra, Itália, Paraguai, Peru, Portugal, Reino Unido, Suíça e Uruguai. Após isso, foram obtidas todas as sequências desses países disponíveis no *Los Alamos HIV Database* para os subtipos B, C e F1, com fragmento de tamanho entre 2262 e 3254, selecionando o segmento gênico da *pol*. Foi selecionada apenas uma sequência por paciente e excluídas sequências problemáticas.

Após a obtenção das sequências externas, utilizou-se a ferramenta Blast+ (National Center for Biotechnology Information - NCBI) para, a partir de um Blastn (National Center for Biotechnology Information - NCBI) (utilizando-se sequências nucleotídicas), identificar as 100 sequências mais similares para cada sequência da RENAGENO. Isto foi realizado separadamente para os subtipos B, C e F1, onde sequências repetidas (aquelas que foram consideradas similares a mais de uma sequência RENAGENO) foram excluídas. Devido ao fato de muitos estudos não fornecerem a classificação até o nível de CRF (forma recombinante circulante), sequências externas para o dataset CRF31 da RENAGENO-SC foram selecionados fazendo-se um Blastn (National Center for Biotechnology Information - NCBI) para cada sequência CRF31 individualmente. Novamente foram excluídas sequências problemáticas e repetidas. Desta forma, o *dataset* de cada subtipo contem as sequências da RENAGENO-SC mais as 100 sequências mais similares dentre as sequências externas selecionadas e excluindo-se as repetidas. Cada *dataset* de sequências foi então alinhado conforme explicado no item 4.1. O tamanho total dos alinhamentos foi de 942 nucleotídeos para o subtipo C, 883 nucleotídeos para o subtipo B e 909 nucleotídeos para o HIV-1F1 e CRF31_BC.

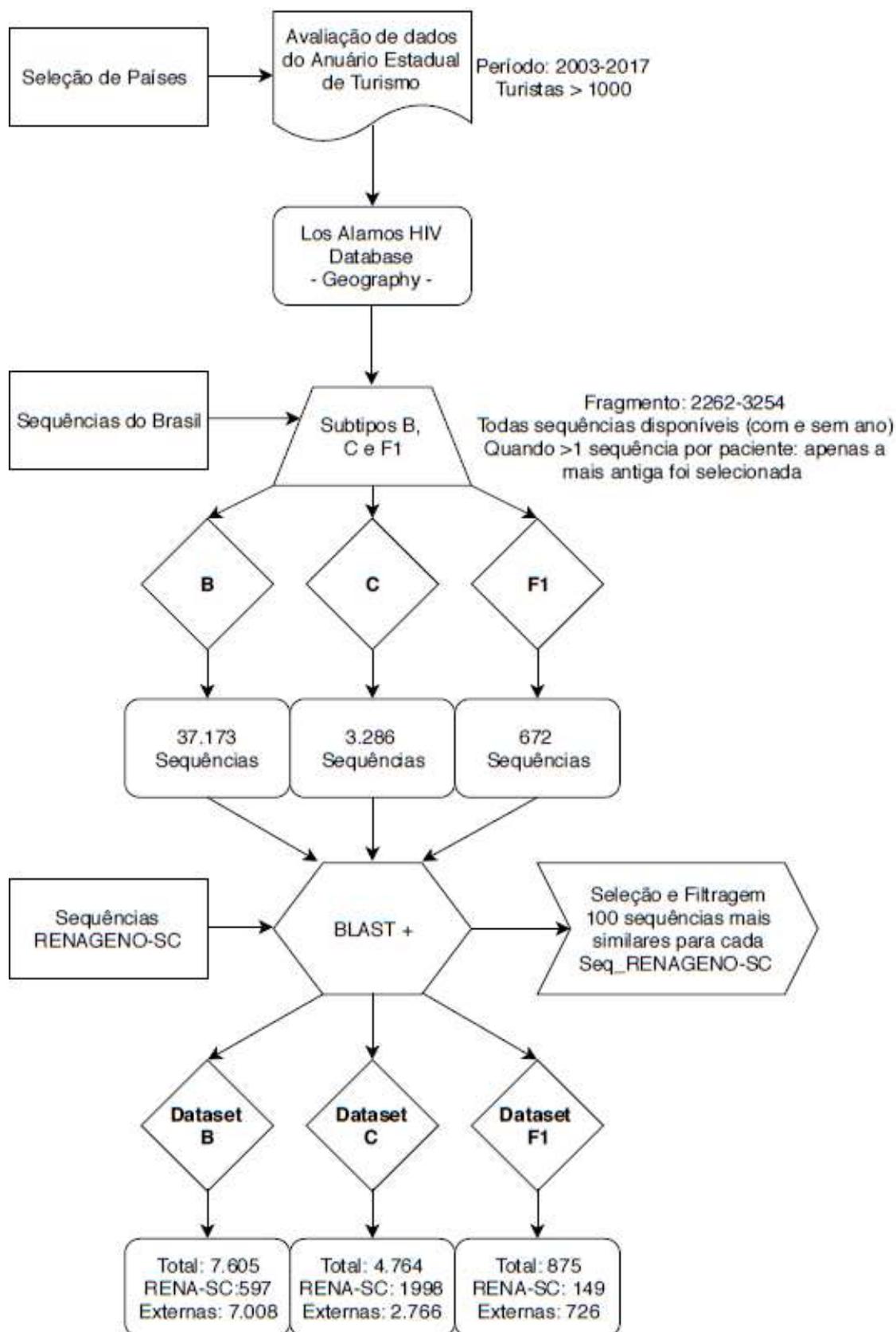


Figura 11: Fluxograma de abordagens utilizadas na seleção de sequências externas ao dataset RENAGENO-SC.

7.3 Identificação das redes de transmissão

Para identificar as redes de transmissão foi realizada uma análise filogenética por máxima verossimilhança (ML) no programa IQ-TREE (Nguyen et al., 2015), utilizando-se a plataforma CIPRES *Science Gateway* (Miller et al., 2010). O melhor modelo de substituição nucleotídica foi testado antes da inferência filogenética pelo método ModelFinder (Kalvaanamoorthy et al., 2017), implementado no IQ-TREE. Como medida de suporte dos clados utilizou-se inicialmente o método de ultrafast bootstrap com 1000 replicatas. Posteriormente, essas 1000 replicatas foram utilizadas para recalculá-las o suporte dos ramos pelo método de transfer bootstrap expectation (TBE) (Lemoine et al., 2018).

O TBE (Lemoine et al., 2018) é um método novo desenvolvido para determinar suportes de ramo. Como vantagem, esse método apresenta uma capacidade aumentada em inferir melhor os suportes de ramos envolvendo nós profundos na árvore filogenética, permitindo então detectar redes de transmissão maiores. Para testar essa condição, foi realizado um teste com três distâncias genéticas (1,5, 4,5 e 8,0) e três opções de suporte de ramo: ultrafast bootstrap, TBE e bootstrap convencional. Essa testagem foi realizada para três *datasets*: C, F1 e CRF31_BC. O *dataset* do subtipo B não foi utilizado para realizar essa etapa em vista de seu tamanho e tempo computacional requerido para refazer as inferências nas três condições de suporte. Como resultado, o TBE mostrou-se capaz de reduzir o número total de clados identificados e aumentando o número de indivíduos clusterizados, ou seja, os clados determinados apresentaram um tamanho maior (Suplementar 1 e 2). Em virtude disso, as árvores de máxima verossimilhança com modificação de suporte para TBE foram selecionadas para as análises posteriores.

A partir da árvore filogenética com suporte de ramos calculado por TBE, os *clusters* de transmissão foram identificados utilizando o programa Cluster Picker (Ragonnet-Cronin et al., 2013) com um limite mínimo de suporte de TBE ≥ 90 . A escolha da distância genética máxima intra-clado foi realizada com base no cálculo da taxa evolutiva dos subtipos B, C, F1 e CRF31 no Brasil. Estudos anteriores estimaram taxas evolutivas de no máximo 0,0022 substituições/sítio/ano para o subtipo B (Mir et al., 2015), 0,0024 para o subtipo

C (Delatorre et al., 2013), 0,0018 para o subtipo F1 (Bello et al., 2012) e 0,0021 para a CRF31 (Passaes et al., 2009). Devido à predominância de pacientes cronicamente infectados no banco da RENAGENO-SC, decidiu-se utilizar um ponto de corte de distância genética pouco restrigente e que possibilite identificar possíveis transmissões que ocorreram ao longo dos últimos 10 anos (tempo aproximado que um indivíduo leva para progredir para aids na ausência de tratamento). Seguindo esta lógica, duas cepas virais podem divergir do seu ancestral comum durante 10 anos, resultando numa divergência acumulada de 20 anos de uma à outra. Desta forma, a taxa evolutiva de cada subtipo foi multiplicada por um tempo máximo de divergência de 20 anos, resultando em distâncias genéticas máximas de 4,4% para o subtipo B, 4,8% para o subtipo C, 3,6% e 4,2% para os subtipos F1 e CRF31_BC respectivamente.

7.4 Caracterização das redes de transmissão

Para cada rede de transmissão identificada foram analisados diversos parâmetros, tais como o número total de redes (a partir daqui também referidos simplesmente como *clusters*) identificados para cada subtipo, o número total de *clusters* que contém pelo menos uma sequência da RENAGENO-SC (RENA *cluster*), o número total de indivíduos clusterizados e a composição entre sequências da RENAGENO e indivíduos externos. Foi determinada também a porcentagem de clusterização para cada subtipo, definida como a razão entre o número de sequências RENAGENO-SC agrupadas e o número total de sequências RENAGENO-SC.

Para identificar a abrangência dos *clusters* RENAGENO-SC identificados foi procedido o cálculo da distância geográfica média entre os indivíduos constituintes do agrupamento de transmissão e observado se essa distância varia entre os subtipos avaliados. Como forma de avaliar essa abrangência territorial foi utilizada a técnica de matriz de distância pelo método elipsóide de Vincenty.

A presença de infecção por HIV-1 em indivíduos com idade igual ou inferior a quatorze anos é um indicativo de possível transmissão vertical. Buscando avaliar a possibilidade de transmissão materno-fetal na população

de estudo, *clusters* de transmissão contendo indivíduos com idade menor ou igual a quatorze anos foram selecionados para uma análise mais minuciosa. Foram selecionados pares de transmissão formados por esses indivíduos e mulheres com idade superior a 15 anos. Como critério de inclusão desses possíveis pares de transmissão foi calculada a diferença de idade entre a possível mãe e a criança, como forma de padronizar o fato das sequências terem sido coletadas em anos distintos. Quando a diferença era inferior a 15 anos, o par foi excluído. As características demográficas da população das possíveis mães envolvidas em eventos de transmissão vertical foi então comparado com as outras mulheres presentes no dataset RENAGENO-SC.

Por fim, ainda estudamos em detalhes clusters formados exclusivamente por homens. Estes clusters foram considerados como possíveis redes de transmissão HSH e as características epidemiológicas desses indivíduos possivelmente foram comparadas com as características obtidas de homens não clusterizados exclusivamente com homens. Foram comparadas também a distância geográfica entre os membros dos agrupamentos, visando determinar a abrangência dessas redes de transmissão em relação a *clusters* de transmissão heterossexual.

7.5 Análises Estatísticas

Para avaliar variáveis qualitativas tais como subtipo viral, gênero, etnia, ano de coleta, escolaridade e clusterização foram utilizados o teste de qui-quadrado ou Odds-Ratio. Para variáveis quantitativas como distância entre os locais de residências de membros constituintes de um cluster de transmissão, foi utilizado teste t-student para amostras independentes ou teste não-paramétrico de Mann-Whitney e teste de correlação de Pearson. Para identificar tendências temporais como o aumento da prevalência dos subtipos durante o período de estudo e aumento de clusterização devido a expansão do tamanho amostral foi utilizado o teste de regressão linear. As análises estatísticas foram executadas no programa R (disponível em: <http://R-project.org/>).

8. RESULTADOS

8.1 Análise comparativa das características epidemiológicas associadas ao agrupamento

Das 3.070 sequências avaliadas, 2.856 (93,0%) compreendiam sequências classificadas nos subtipos B, C, F1 e CRF31_BC do HIV-1, sendo estas então submetidas às análises de investigação e caracterização de redes de transmissão (Tabela 2). Após a identificação de *clusters* através da ferramenta ClusterPicker foi possível observar que 943 (33,0%) sequências da RENAGENO-SC estavam contidas em alguma rede de transmissão. Dos indivíduos que estavam contidos nesses *clusters*, 47,9% eram mulheres. Ao analisar as características epidemiológicas dos indivíduos clusterizados em comparação aos que não clusterizaram, foi possível observar que indivíduos entre 15 e 24 anos, 41 e 59 anos e com idade superior a 59, apresentaram chances 40 a 50% menor de estarem contidos em algum agrupamento ($p=0,01$, $p<0,001$ e $p=0,01$, respectivamente) quando comparados aos indivíduos com 25 a 40 anos (que foram utilizados como referência na análise de Odds Ratio).

O ano de coleta das sequências também mostrou-se um importante fator a impactar o agrupamento. Indivíduos amostrados entre os anos de 2014 e 2017 apresentaram 1,4 vezes mais chances de serem encontrados em um cluster de transmissão (Tabela 2). Devido a esse aumento da probabilidade de clusterização observado a partir de 2014 e ao fato de neste ano o Ministério da Saúde ter mudado o protocolo de início da TARV, investigamos com mais atenção a correlação entre amostragem e o número de sequências contidas em *clusters*. Através do teste de correlação de Pearson foi possível observar que há uma forte correlação ($p < 0,001$, $r=0,95$) entre o número de sequências amostradas e o número de sequências contidas em clusters de transmissão por ano (Figura 12).

Outro fator associado ao agrupamento foi o subtipo viral, uma vez que o subtipo C foi observado como sendo muito mais associado com a formação de redes de transmissão. Quando comparado ao HIV-1C, indivíduos infectados

pelos subtipos B, F1 e CRF31_BC apresentaram uma redução de 40, 90 e 50%, respectivamente, na chance de clusterização.

O agrupamento de sequências não apresentou influência de gênero ($p=0,28$), visto que homens e mulheres apresentaram chances similares de estarem contidos em um cluster de transmissão. Além disso, não foram encontradas associações estatísticas significativas entre a formação de *clusters* de transmissão e etnia ($p=0,09$) e escolaridade (p -values entre 0,34 e 0,68). Embora a escolaridade não tenha apresentado diferenças estatísticas, indivíduos com 1 a 3 anos de acesso a educação apresentaram um indício ($p=0,06$) de redução de 30% na chance de clusterizarem ao comparado com indivíduos com 4-7 anos de escolaridade.

Tabela 2: Dados epidemiológicos associados com o agrupamento geral em pacientes da RENAGENO-SC.

Clusterização				
	Sequências Agrupadas n=943 (33.0)	Sequências Não Agrupadas n= 1913 (67.0)	OR (95%IC)	P value
Gênero				
Feminino	452 (47,9)	944 (49,3)	(Ref)	(Ref)
Masculino	491 (52,1)	969 (50,7)	1,1 (0,9 – 1,3)	0,28
Etnia				
Não-Caucasiano	104 (28,8)	257 (71,2)	0,7 (0,4 – 1,0)	0,09
Caucasiano	678 (32,4)	1412 (67,6)	(Ref)	(Ref)
Anos de Escolaridade				
Nenhum	26 (35,6)	47 (64,4)	1,1 (0,6 – 2,1)	0,61
1-3 anos	64 (24,5)	197 (75,5)	0,7 (0,5 – 1,0)	0,06
4-7 anos	302 (31,6)	652 (68,4)	(Ref)	(Ref)
8-11 anos	227 (35,4)	415 (64,6)	1,1 (0,8 – 1,4)	0,34
12 ou mais anos	68 (31,5)	148 (68,5)	0,9 (0,7 – 1,3)	0,68
Idade				
<14 anos	93 (32,9)	190 (67,1)	0,8 (0,5 – 1,2)	0,23
15-24 anos	75 (34,5)	142 (65,4)	0,6 (0,4 – 0,9)	0,01
25-40 anos	428 (39,7)	651 (60,3)	(Ref)	(Ref)
41-59 anos	319 (27,4)	844 (72,6)	0,6 (0,5 – 0,7)	< 0,001
>59 anos	28 (24,6)	86 (75,4)	0,5 (0,3 – 0,9)	0,01
Ano de Coleta				
2008-2013	398 (42,2)	966 (50,5)	(Ref)	(Ref)
2014-2017	545 (57,8)	947 (49,5)	1,4 (1,1 – 1,7)	< 0,001
Subtipos				
B	152 (25,5)	445 (74,5)	0,6 (0,4 – 0,7)	< 0,001
C	737 (36,9)	1261 (63,1)	(Ref)	(Ref)
F1	18 (12,1)	131 (87,9)	0,1 (0,1 – 0,3)	< 0,001
CRF31	36 (32,2)	76 (67,8)	0,5 (0,3 – 0,9)	0,03

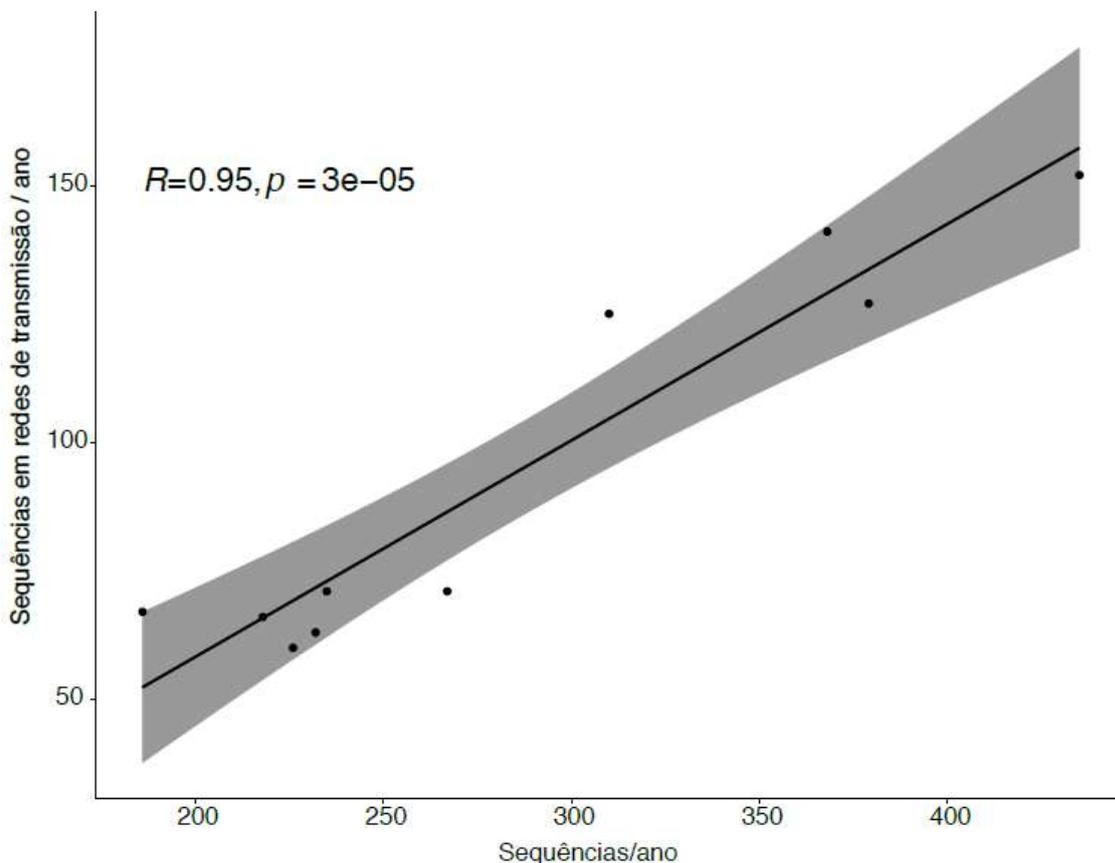


Figura 12: Correlação entre o número de sequências amostradas por ano e o número de sequências agrupadas no *dataset* RENAGENO-SC.

O aumento da densidade amostral de sequências coletadas representou um aumento no número de sequências identificadas em redes de transmissão por ano, sendo um indicativo de que quanto mais pacientes são amostrados, melhor é a capacidade de identificação e reconstrução dessas relações de transmissão. É esperado que após as mudanças na política de acesso e tratamento universal de todo paciente a partir do diagnóstico o banco de dados da RENAGENO forneça informações suficientes para identificar e reconstruir as redes de transmissão atuantes.

8.2 Análise comparativa das características epidemiológicas associadas à clusterização nos diferentes subtipos

A identificação das redes de transmissão foi realizada com distâncias genéticas máximas determinadas pelo cálculo de taxa evolutiva referente a cada subtipo e um período máximo de divergência gênica de 20 anos entre duas cepas virais (Tabela 3). A partir disso foram encontrados 94 *clusters* para o subtipo B e 341 para o subtipo C, sendo os dois subtipos com maior número de *clusters* envolvendo indivíduos da RENAGENO-SC. Para os subtipos F1 e CRF31_BC foram encontrados 10 e 22 clusteres, respectivamente.

Quando a composição de *clusters* foi analisada observou-se que o subtipo C apresentou a maior taxa de clusterização, sendo que 36,9% das sequências HIV-1C no banco RENAGENO-SC estavam presentes em alguma rede de transmissão. O subtipo B apresentou uma porcentagem de clusterização de 25,4% e o maior número total de *clusters* observado para este subtipo deve-se ao grande número de sequências externas selecionadas. Entretanto apenas 9,2% do total de *clusters* continha uma sequência RENAGENO-SC. A porcentagem de clusterização da CRF31 (33,9) foi similar à encontrado para o subtipo C; entretanto o HIV-1F1 apresentou a menor taxa de clusterização em comparação aos outros subtipos avaliados, onde apenas 12,1% dos indivíduos infectados por essa forma viral foram classificados como participando de uma rede de transmissão.

Para todos os subtipos foi encontrado um padrão de detecção de *clusters* pequenos, envolvendo principalmente pares de transmissão, exceto para o HIV-1C. Além de apresentar o maior número de *clusters* envolvendo indivíduos da RENAGENO-SC, o subtipo C apresentou também maior diversidade quanto ao tamanho dos agrupamentos encontrados (Tabela 3 e Figura 13), incluindo um cluster contendo 10 indivíduos (Figura 14).

Tabela 3: Análise comparativa das redes de transmissão identificadas para os subtipos B, C, F1 e CRF31_BC do HIV-1.

	Subtipos			
	B	C	F1	CRF31_BC
Análise de Clusters				
Total de <i>clusters</i>	1020	700	70	191
Total de RENAGENO-SC <i>clusters</i> *	94 (9,2)	341(48,7)	10 (14,3)	22 (11,5)
Total de indivíduos nos RENAGENO-SC <i>clusters</i>	217	851	22	57
Indivíduos RENAGENO-SC agrupados	152 (70,0)	738 (86,7)	18 (81,8)	38 (66,7)
Porcentagem de agrupamento**	25,4	36,9	12,1	33,9
Tamanho dos <i>clusters</i> ***				
2 indivíduos	73 (77,6)	237 (69,5)	8 (80,0)	13 (59,1)
3 indivíduos	14 (14,9)	68 (20,0)	1 (10,0)	5 (22,7)
4 indivíduos	6 (6,4)	20 (5,9)	1 (10,0)	4 (18,2)
5 indivíduos	1 (1,1)	8 (2,3)	0 (0,0)	0 (0,0)
6 indivíduos	0 (0,0)	6 (1,7)	0 (0,0)	0 (0,0)
7 indivíduos	0 (0,0)	1 (0,3)	0 (0,0)	0 (0,0)
10 indivíduos	0 (0,0)	1 (0,3)	0 (0,0)	0 (0,0)

* *clusters* contendo pelo menos uma sequência RENAGENO-SC; **razão entre o número de sequências analisadas e o número de sequências agrupadas; ***considerando apenas RENAGENO-SC *clusters*

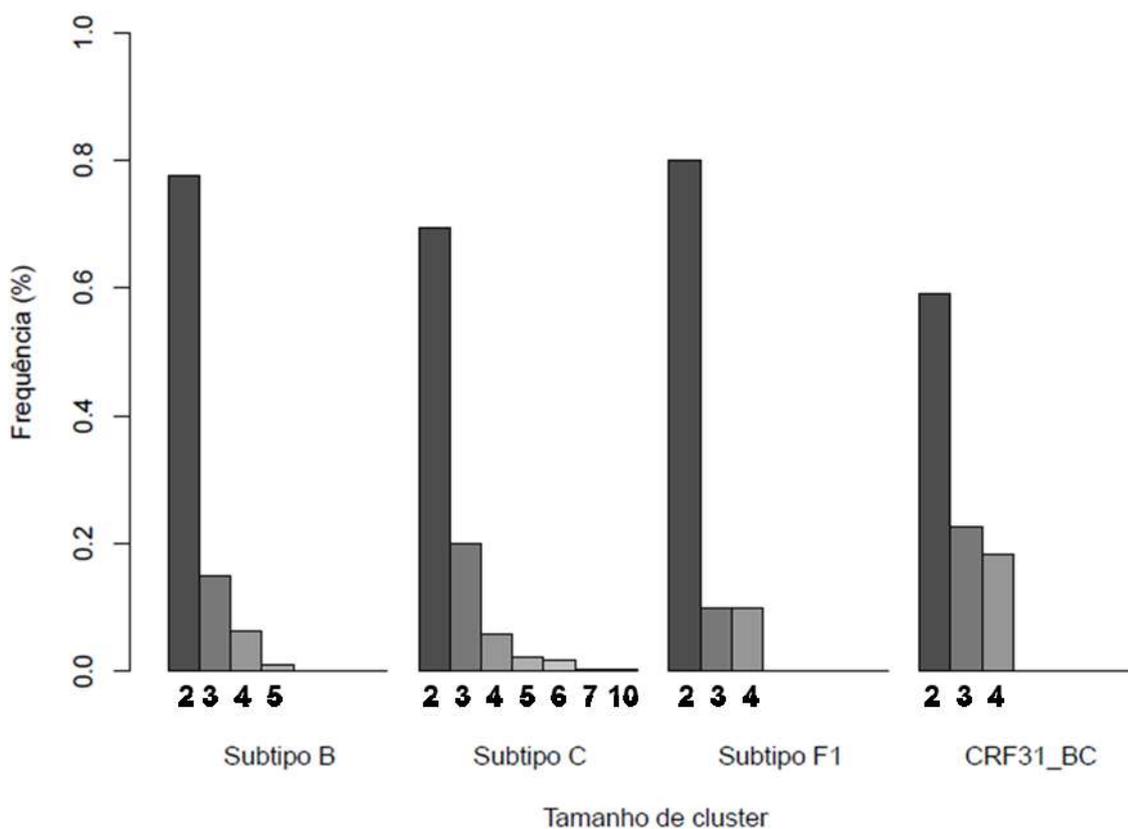


Figura 13: Distribuição da frequência obtida e tamanhos de *clusters* de transmissão para os subtipos B, C, F1 e CRF31_BC do HIV-1.

Nosso estudo identificou, para os quatro subtipos avaliados, uma elevada proporção de *clusters* formados por apenas dois indivíduos (também denominados pares de transmissão) e agrupamentos pequenos. Esse achado pode representar um viés de amostragem onde, em vista da impossibilidade de amostrar a totalidade da população infectada, a reconstrução das redes de transmissão torna-se prejudicada pela falta de indivíduos considerados elos epidemiológicos.

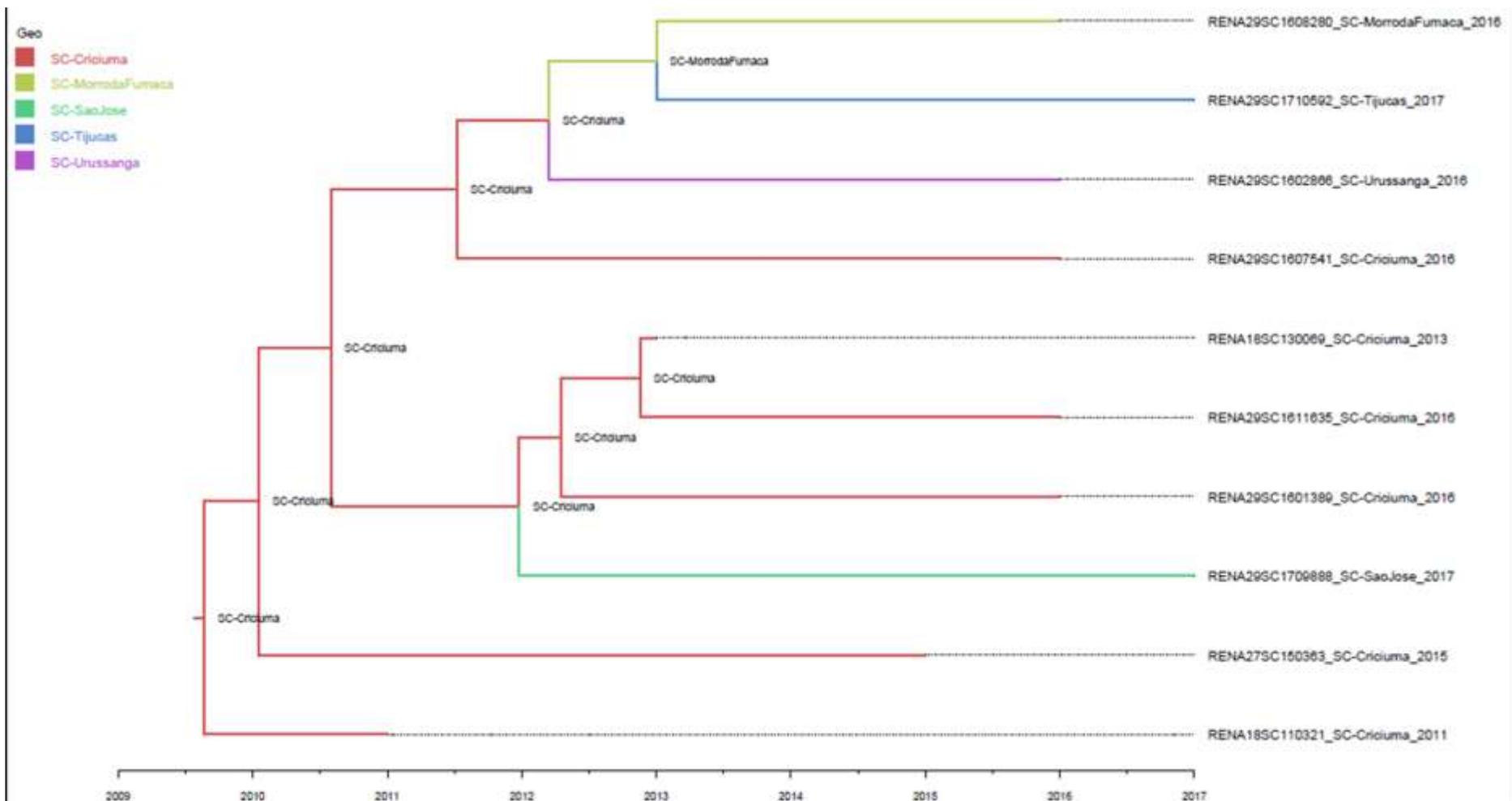


Figura 14: Representação filogenética do maior *cluster* de transmissão do HIV-1C, identificado pela ferramenta Cluster Picker, no estado de Santa Catarina. Amostragem de 10 indivíduos potencialmente relacionados epidemiologicamente por eventos de transmissão.

Em vermelho estão representados os nodos e ramos associados ao município de Criciúma, em verde-oliva ao município de Morro da Fumaça, verde a São José, azul e roxo aos municípios de Tijucas e Urussanga, respectivamente. É possível observar, a partir do nodo mais ancestral da filogenia, que as redes de transmissão atuantes cidade de Criciúma podem estar associadas à disseminação de subtipo C para outros municípios e regiões intermediária do estado, sendo origem da transmissão para São José, Urussanga e Morro da Fumaça. Além disso, a partir da filogenia é possível inferir também que possivelmente houve transmissão envolvendo Morro da Fumaça para Tijucas. Contudo, mais estudos são necessários para melhor compreender a dinâmica de transmissão atuante nessa rede.

Em vista dos indícios de haver diferenças marcantes entre as chances de clusterização entre os subtipos amostrados, foram analisados os fatores epidemiológicos associados aos HIV-1B, C, F1 e CRF31_BC que poderiam impactar nas chances de clusterização. Quando avaliadas separadamente as características epidemiológicas de indivíduos clusterizados e não clusterizados para os subtipos avaliados não houve acréscimo de informações quanto à clusterização. Aparentemente apenas a idade desses indivíduos impactou no agrupamento deles em *clusters* (Tabela 4).

Tabela 4: Dados epidemiológicos associados as diferenças em agrupamentos observados para cada subtipo avaliado.

Clusterização	Subtipos															
	B				C				F1				CRF31			
	Sim n=152 (25,5)	Não n=445 (74,5)	OR (95% IC)	P value	Sim n=737 (36,9)	Não n=1261 (63,1)	OR (95% IC)	P value	Sim n=18 (12,1)	Não n=131 (87,9)	OR (95% IC)	P value	Sim n=36 (32,2)	Não n=76 (67,8)	OR (95%IC)	P value
Gênero																
Feminino	51 (22,0)	181 (78,0)	(Ref)	(Ref)	374 (35,9)	667 (64,1)	(Ref)	(Ref)	7 (10,6)	60 (89,6)	(Ref)	(Ref)	20 (35,7)	36 (64,3)	(Ref)	(Ref)
Masculino	101 (27,7)	264 (72,3)	1,4 (0,8 – 2,2)	0,17	363 (37,9)	594 (62,1)	1,1 (0,9 – 1,4)	0,37	11 (13,4)	71 (86,6)	0,7 (0,1 – 3,3)	0,62	16 (28,6)	40 (71,4)	0,6 (0,2 – 1,9)	0,42
Etnia																
Não-caucasiano	16 (20,0)	64 (80,0)	0,7 (0,3 – 1,3)	0,26	81 (31,4)	177 (68,6)	0,7 (0,5 – 1,0)	0,08	2 (20,0)	8 (80,0)	6,9 (0,3 - 160)	0,19	5 (38,5)	8 (61,5)	1,1 (0,2 – 4,6)	0,85
Caucasiano	116 (25,7)	335 (74,3)	(Ref)	(Ref)	526 (36,6)	909 (63,3)	(Ref)	(Ref)	10 (8,3)	111 (91,7)	(Ref)	(Ref)	26 (31,3)	57 (68,7)	(Ref)	(Ref)
Anos de Escolaridade																
Nenhum	5 (41,7)	7 (58,3)	1,6 (0,3 – 6,8)	0,52	20 (36,4)	35 (63,4)	1,0 (0,5 – 2,1)	0,87	1 (33,3)	2 (66,7)	103 (-)	0,12	0 (0,0)	3 (100,0)	1,1 (0,6 – 2,0)	0,99
1-3 anos	7 (14,6)	41 (85,4)	0,5 (0,2 – 1,1)	0,11	56 (29,5)	134 (70,5)	0,8 (0,6 – 1,2)	0,28	0 (0,0)	16 (100,0)	0,0 (-)	0,99	1 (14,3)	6 (85,7)	0,7 (0,5 – 1,0)	0,49
4-7 anos	47 (27,5)	124 (72,5)	(Ref)	(Ref)	242 (35,1)	447 (64,9)	(Ref)	(Ref)	1 (2,1)	47 (97,9)	(Ref)	(Ref)	12 (26,1)	34 (73,9)	(Ref)	(Ref)
8-11 anos	39 (26,2)	110 (73,8)	0,8 (0,5 – 1,3)	0,36	174 (40,5)	255 (59,5)	1,1 (0,9 – 1,5)	0,36	8 (18,2)	36 (81,8)	13 (-)	0,05	6 (30,0)	14 (70,0)	1,1 (0,8 – 1,3)	0,45
12 ou mais anos	22 (25,0)	66 (75,0)	0,7 (0,4 – 1,3)	0,36	44 (39,3)	68 (60,7)	1,0 (0,7 – 1,5)	0,92	1 (11,1)	8 (88,9)	2,9 (-)	0,58	1 (14,3)	6 (85,7)	0,8 (0,6 – 1,2)	0,55
Idade																
<14 anos	9 (21,4)	33 (78,6)	0,6 (0,2 – 1,7)	0,41	80 (36,2)	141 (63,8)	0,8 (0,5 – 1,3)	0,37	3 (30,0)	7 (70,0)	0,3 (-)	0,67	1 (10,0)	9 (90,0)	2,9 (0,1 – 92,6)	0,49
15-24 anos	10 (34,5)	19 (65,5)	0,8 (0,3 – 2,2)	0,74	63 (37,5)	105 (62,5)	0,6 (0,4 – 0,9)	0,03	2 (20,0)	8 (80,0)	3,0 (0,0 - 152)	0,57	0 (0,0)	10 (100,0)	0,0 (-)	0,99
25-40 anos	80 (30,7)	181 (69,3)	(Ref)	(Ref)	323 (45,3)	390 (54,7)	(Ref)	(Ref)	10 (16,7)	50 (83,3)	(Ref)	(Ref)	15 (33,3)	30 (66,7)	(Ref)	(Ref)
41-59 anos	53 (21,1)	198 (78,9)	0,6 (0,4 – 0,9)	0,02	246 (30,5)	561 (69,5)	0,5 (0,4 – 0,7)	< 0,001	2 (3,1)	62 (96,9)	0,07 (0,0 – 0,5)	0,04	18 (43,9)	23 (56,1)	2,1 (0,6 – 7,9)	0,24
>59 anos	0 (0,0)	14 (100,0)	0,6 (0,3 – 0,9)	0,98	25 (28,1)	64 (71,9)	0,6 (0,3 – 1,0)	0,08	1 (20,0)	4 (80,0)	0,0 (-)	0,99	2 (33,3)	4 (66,7)	1,1 (0,0 – 11,2)	0,95

8.3 Análise da composição geográfica das redes de transmissão e agrupamento externo

Com a finalidade de avaliar a extensão geográfica das redes envolvendo pacientes da RENAGENO-SC, *datasets* de trabalho foram montados para cada subtipo contendo sequências internacionais provenientes dos países que mais enviam turistas para o estado de Santa Catarina e de outros estados federativos brasileiros disponíveis no *Los Alamos HIV Database* (Tabela 5).

O *dataset* do subtipo B do HIV-1 apresentou o maior volume de dados (7.605 sequências) avaliados, sendo 92,2% do total de sequências constituintes provenientes de outros países e outros estados. Já o *dataset* para o HIV-1C contabilizou 4.764 sequências, sendo 1.998 (42,0%) de sequências provenientes de pacientes RENAGENO-SC e 58% de sequências externas. A redução drástica do número de sequências externas a Santa Catarina amostradas nesse *dataset* em comparação ao subtipo B está relacionado a baixa prevalência de subtipo C em outros estados federativos brasileiros, exceto os da região Sul, e baixa prevalência em países amostrados, nos quais o subtipo B é responsável, majoritariamente, pelas infecções. Os *datasets* formados para o subtipo F1 e a forma recombinante CRF31 foram menores em relação aos subtipos B e C, nos quais 17 e 10,1% das sequências eram provenientes de pacientes de Santa Catarina.

Ao analisar a composição dos *clusters* encontrados (Tabela 5), subtipo C e F1 apresentaram um maior proporção de *clusters* formados exclusivamente por sequências da RENAGENO-SC (77,7 e 70,0%, respectivamente), enquanto, para os subtipos B e a forma recombinante CRF31_BC foram encontrados maior proporção de *clusters* mistos (53,2 e 59,1%, respectivamente), formados por sequências provenientes do estado de Santa Catarina com sequências externas.

A origem das sequências externas agrupadas também foi avaliada, sendo que a partir dos dados obtidos para cada subtipo foi possível inferir que o padrão de clusterização com sequências internacionais é diferente para subtipos B e C. Sequências do HIV-1B associadas aos pacientes RENAGENO-

SC apresentaram relações epidemiológicas com sequências provenientes do Reino Unido, Chile e França. Já para o subtipo C foram descritas clusterizações envolvendo Argentina, Alemanha, Espanha e Reino Unido e Estados Unidos. O subtipo F1 apresentou apenas duas interconexões epidemiológicas externas (Estados Unidos e Espanha). Embora no Capítulo 1 deste trabalho tenha sido descrita associação entre elevada prevalência de HIV-1F1 e região intermediária fronteira com a Argentina, não foram encontradas interconexões epidemiológicas externas com este país. Não foram encontradas associações internacionais para a forma recombinante CRF31, evidenciando sua circulação local.

Embora o padrão de clusterização internacional seja diferente, subtipos B e C apresentam um padrão similar de clusterização quando analisados os estados brasileiros, uma vez que tanto HIV-1B e C apresentam forte interconexão com os estados da região sul do país. Majoritariamente, as sequências B e C da RENAGENO-SC agrupam-se com sequências provenientes dos estados do Paraná (20,7 e 21,2%, para subtipos B e C) e Rio Grande do Sul (17,2% para ambos subtipos). Além disso, foram encontradas associações entre a epidemia de Santa Catarina e o estado de São Paulo, envolvendo o subtipo B (15,5%) e Rio de Janeiro (13,2%) para o HIV-1C. Estados com associação inferior a 10% foram incluídos em um grupo denominado “outros”, que contabilizaram 20,7 e 19,1% das relações epidemiológicas encontradas pra B e C, respectivamente. O subtipo F1 foi o que apresentou menores relações epidemiológicas interestaduais, envolvendo apenas três associações com Rio de Janeiro, Paraná e Bahia (33,33% cada). Simultaneamente, a CRF31_BC apresentou 35,7% de relações com o Rio Grande do Sul, estando estas associadas possivelmente a elevada prevalência dessa forma recombinante nesse estado. As sequências brasileiras que não possuíam dados associados a localidade de origem (estado) foram classificadas como BR-NA.

Tabela 5: Caracterização da composição geográfica das redes de transmissão para os subtipos B, C, F1 e CRF31_BC.

	Subtipos			
	B	C	F1	CRF31_BC
Datasets				
Total de sequências	7605	4764	875	1109
Total de sequências RENAGENO-SC	597 (7,9)	1998 (42,0)	149 (17,0)	112 (10,1)
Total de sequências externas	7008 (92,2)	2766 (58,0)	726 (83,0)	997 (89,9)
Clusterização				
RENAGENO-SC <i>clusters</i>	94	341	10	22
Sequências RENAGENO-SC agrupadas*	152 (25,4)	738 (36,9)	18 (12,1)	38 (33,9)
Sequências externas nos RENAGENO-SC <i>clusters</i> **	65 (0,9)	113 (4,1)	5 (0,7)	19 (1,9)
Composição do Cluster				
<i>Clusters</i> exclusivos de sequências RENAGENO-SC	44 (46,8)	265 (77,7)	7 (70,0)	9 (40,9)
<i>Clusters</i> mistos	50 (53,2)	76 (22,3)	3 (30,0)	13 (59,1)
Origem das sequências externas nos RENAGENO-SC <i>clusters</i>				
Internacional	n=7	n=14	n=2	n=0
Países ***	GRB (57,1) CHI (28,6) FRA (14,3)	ARG (28,6) DEU (28,6) ESP (21,4) GRB (14,3) USA (7,1)	USA (50,0) ESP (50,0)	-
Brasil	n=58	n=99	n=3	n=14
Estados Brasileiros ***	BR-NA (25,9) PR (20,7) RS (17,2) SP (15,5) Outros (20,7)	BR-NA (29,3) PR (21,2) RS (17,2) RJ (13,2) Outros (19,1)	RJ (33,33) PR (33,33) BA (33,33)	BR-NA (50,0) RS (35,7) Outros (14,3)

* porcentagem sobre o total de sequências RENAGENO-SC; ** porcentagem sobre total de sequências externas; *** porcentagem de sequências externas nos RENAGENO-SC *clusters*. ARG: Argentina, DEU: Alemanha, GRB: Reino Unido, ESP: Espanha, CHI: Chile, FRA: França, USA: Estados Unidos. BR-NA sequências provenientes do Brasil sem informação de localidade de origem, PR: Paraná, RS: Rio Grande do Sul, SP: São Paulo, RJ: Rio de Janeiro, BA: Bahia.

8.4 Análise geoespacial da distribuição dos *clusters* de transmissão em Santa Catarina

Buscando avaliar a abrangência dos *clusters* de transmissão pelo território de Santa Catarina, foi calculada uma matriz de distância entre as cidades de origem dos membros dos *clusters* formados exclusivamente por indivíduos RENAGENO-SC. A distância entre as cidades foi estimada através do método elipsóide de Vincenty que basicamente calcula a distância em linha reta entre os pontos corrigido pela curvatura da terra.

Foram descritos 44 *clusters* exclusivamente RENAGENO-SC para o subtipo B, 265 para o subtipo C, 7 e 9 para o HIV-1F1 e CRF31_BC, respectivamente (Tabela 6). Além de apresentar o maior número de *clusters*, o subtipo C também apresenta a maior mediana de distância (21 km) entre as cidades de residência dos membros compondo cada *cluster*, sendo este valor estatisticamente maior que as medianas de distância dos outros subtipos (Figura 14). Isso pode estar relacionado ao fato do subtipo C possuir *clusters* maiores que os outros subtipos, sendo o único a ter *clusters* com mais de cinco indivíduos quando consideradas apenas as sequências RENAGENO-SC. Quanto maior o *cluster*, maior a probabilidade de pessoas de cidades diversas serem incluídas.

Além disso, quando foi avaliada a proporção de *clusters* categorizados por distância média (0, 1-50 e >50km), é possível observar que todos os subtipos possuem majoritariamente *clusters* formados por indivíduos da mesma cidade (0 km de distância). Entretanto, para o HIV-1C a proporção de *clusters* com distâncias maiores de 50km (36,6%) é muito semelhante à proporção de *clusters* intra-municipais, sendo a maior quando comparado aos outros subtipos (Tabela 6). Interessante notar também que o subtipo F1 apresentou quase que exclusivamente *clusters* intra-municipais e 100% dos *clusters* eram pares. Isso reforça a hipótese que *clusters* pequenos tendem a agrupar-se localmente, enquanto redes maiores podem se espalhar territorialmente por uma região mais vasta.

Tabela 6: Caracterização de tamanho e distância entre membros de *clusters* de transmissão constituídos exclusivamente por pacientes RENAGENO-SC.

Clusterização	Subtipos			
	B	C	F1	CRF31
Total de <i>clusters</i>	n = 44	n=265	n=7	n=9
Mediana (km)	3,0	21,0	0	17,0
Tamanho dos <i>clusters</i>				
2 indivíduos	35 (79,6)	200 (75,5)	7 (100,0)	5 (55,6)
3 indivíduos	6 (13,6)	46 (17,3)	0 (0,0)	2 (22,2)
4 indivíduos	3 (6,8)	14 (5,3)	0 (0,0)	2 (22,2)
5 indivíduos	0 (0,0)	2 (0,75)	0 (0,0)	0 (0,0)
6 indivíduos	0 (0,0)	1 (0,38)	0 (0,0)	0 (0,0)
7 indivíduos	0 (0,0)	1 (0,38)	0 (0,0)	0 (0,0)
10 indivíduos	0 (0,0)	1 (0,38)	0 (0,0)	0 (0,0)
Distância				
0 km	22 (50,0)	98 (37,0)	6 (85,7)	2 (22,2)
1-50 km	14 (31,8)	70 (26,4)	1 (14,3)	5 (55,6)
>50 km	8 (18,2)	97 (36,6)	0 (0,0)	2 (22,2)

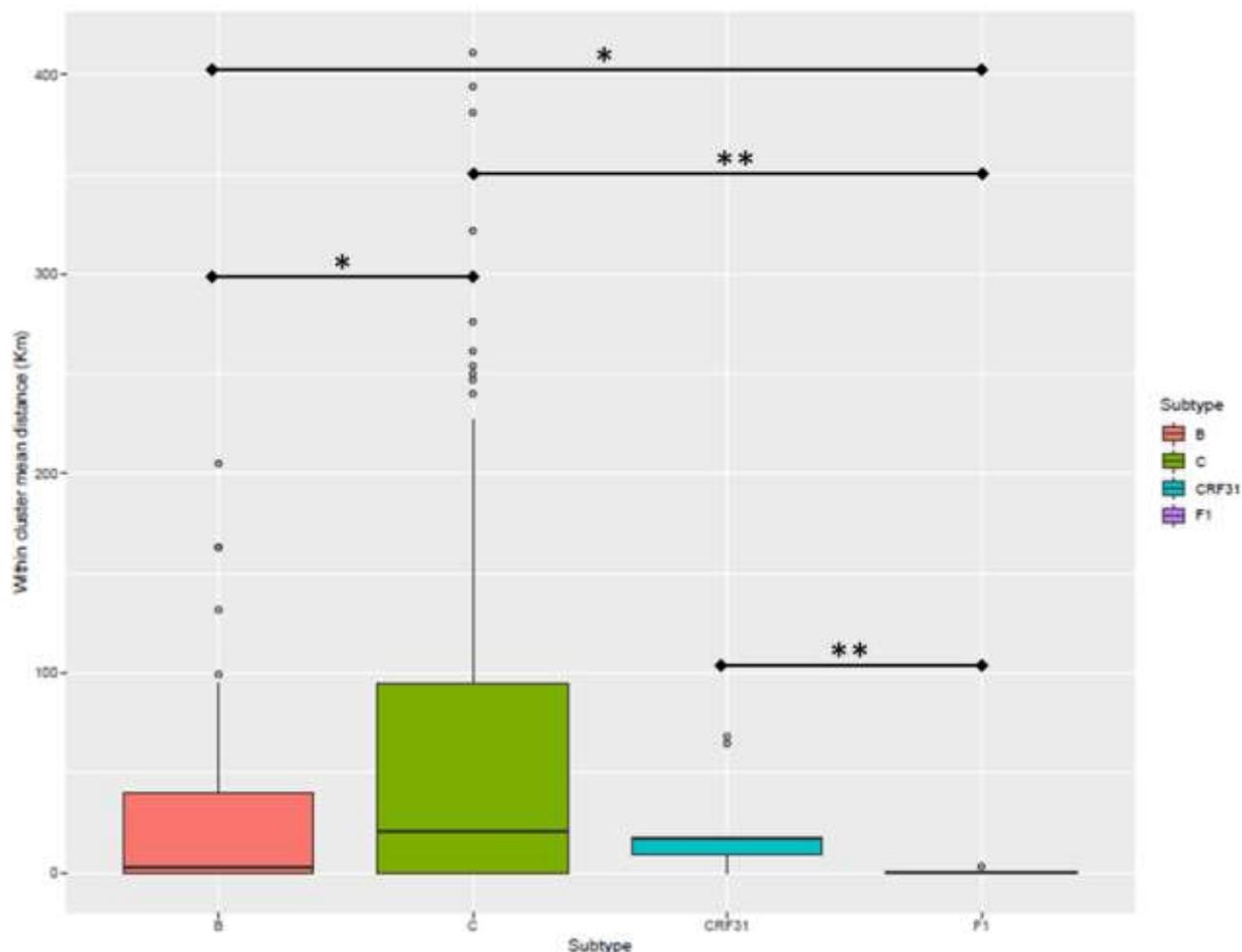


Figura 15: Distribuição das distâncias médias entre indivíduos compondo *clusters* de transmissão exclusivamente RENAGENO-SC.

As barras indicam as diferenças significativa entre as médias de distância obtidas para cada subtipo. * $p=0,05$, ** $p<0,01$, *** $p<0,001$.

O subtipo C do HIV-1 apresentou a maior mediana de distância entre os membros dos agrupamentos identificados. Essa distância mostrou-se superior a encontrada para o HIV-1B ($p=0,05$) e F1 ($p<0,01$), sendo um indicativo de que as redes de transmissão atuantes nesse subtipo são mais extensas, abrangendo um número maior de indivíduos e, conseqüentemente, municípios distintos. A baixa mediana encontrada para o HIV-1B pode ser decorrente ao menor número de clusters avaliados, visto que 53,2% dos agrupamentos possuíam sequências externas, sendo excluídos das análises geoespaciais.

8.5 Análise de agrupamento em indivíduos com idade inferior a 14 anos e indicativo de possível transmissão vertical

A presença de infecção por HIV-1 em indivíduos com idade igual ou inferior a quatorze anos é um indicativo de possível transmissão vertical, uma vez que eles teriam se infectado antes do início da vida sexual. Durante o período de coleta das sequências, foram amostrados 283 (0,09%) indivíduos com idade menor ou igual a 14 anos. Destes, 150 (53,0%) foram do sexo feminino e 133 (47,0%) do sexo masculino, sendo a população majoritariamente caucasiana (86,7%). Quando analisamos mais detalhadamente a idade desses indivíduos, observamos que 36,4% tinham idade entre 12 e 14 anos e 14,0% idade entre 9 e 11. Também observamos que 49 indivíduos (17,3%) tinham idade até 2 anos completos e a faixa etária de 3 a 5 e 6 a 8 anos totalizaram 10,6% e 11,6% dos indivíduos amostrados, respectivamente. O HIV-1C foi o mais presente nesse grupo de indivíduos, sendo responsável por 78,1% das infecções, seguido pelo subtipo B (14,8%). Foi encontrada baixa prevalência de subtipos F1 e CRF31, totalizando 3,5% de casos cada.

Quando o agrupamento desses indivíduos foi analisado (Tabela 7) observou-se que 62,4% dos indivíduos do sexo feminino clusterizaram em comparação a 37,6% no sexo masculino ($p=0,03$). O agrupamento foi associado negativamente a indivíduos de 12 a 14 anos e positivamente em indivíduos com menos de 2 anos de idade ($p=0,003$). Embora o subtipo C tenha sido encontrado em 86% dos indivíduos que agruparam, não foi encontrada nenhuma associação estatística significativa entre subtipo e o agrupamento em indivíduos com idade inferior a 14 anos ($p=0,11$), bem como não foi encontrada associação entre clusterização e etnia ($p=0,86$) e ano de coleta ($p=0,71$) (Tabela 7).

Tabela 7: Agrupamento diferencial de crianças com idade inferior a 14 anos e associação com dados epidemiológicos.

	<14 Agrupados n=93 (32,9)	<14 Não Agrupados n=190 (67,1)	P value
Gênero			
Feminino	53 (62,4)	92 (48,4)	0,03
Masculino	35 (37,6)	98 (51,6)	
Etnia			
Não Caucasiano	10 (15,0)	17 (12,3)	0,86
Caucasiano	57 (85,0)	121 (87,7)	
Idade			
0-2 anos	22 (23,7)	27 (14,2)	0,003
3-5 anos	14 (15,0)	16 (8,4)	
6-8 anos	14 (15,0)	19 (10,0)	
9-11 anos	23 (24,7)	45 (23,7)	
12-14 anos	20 (21,5)	83 (43,7)	
Subtipo			
B	9 (9,7)	33 (17,4)	0,11
C	80 (86,0)	141 (74,2)	
F1	3 (3,2)	7 (3,7)	
CRF31_BC	1 (1,1)	9 (4,7)	
Ano de Coleta			
2008-2009	27 (29,0)	63 (33,2)	0,71
2010-2011	19 (20,4)	47 (24,7)	
2012-2013	20 (21,5)	34 (17,9)	
2014-2015	13 (14,0)	25 (13,2)	
2016-2017	14 (15,1)	21 (11,0)	

Na tentativa de elucidar e identificar possíveis casos de transmissão vertical na população de infectados com menos de 14 anos, foi analisado também as possíveis mães que clusterizaram com esses indivíduos. Foram identificados 19 pares de transmissão entre crianças com menos de 14 anos e mulheres acima de 15 anos. Destes 19 pares de transmissão, 8 pares criança-possível mãe (42,1%) residiam na mesma cidade e 11 (57,9%) em cidades distintas. As mulheres que clusterizaram (Tabela 8) representam 1,5% das

mulheres da população do estudo, contudo não foram encontradas associações estatísticas significativas que pudessem diferenciá-las do restante da população feminina amostrada. Essa análise foi repetida também analisando apenas os 8 pares residentes na mesma cidade, contudo também não foram encontradas associações significativas (dados não mostrados).

Tabela 8: Características epidemiológicas associadas às possíveis mães agrupadas com crianças com idade inferior a 14 anos.

	Possíveis mães transmissoras n=19 (1,5)	Mulheres em geral n=1227 (98,5)	OR (95% IC)	P value
Etnia				
Não-Caucasiano	3 (17,7)	184 (16,9)	0,9 (0,2 – 3,3)	0,93
Caucasiano	14 (82,3)	908 (83,1)	(Ref)	(Ref)
Anos de Escolaridade				
Nenhum	0 (0,0)	8 (0,8)	0,0 (-)	0,99
1-3 anos	2 (13,3)	123 (13,0)	1,0 (0,1 – 4,5)	0,96
4-7 anos	8 (53,4)	449 (47,4)	(Ref)	(Ref)
8-11 anos	3 (20,0)	299 (31,5)	0,7 (0,1 – 2,6)	0,62
12 ou mais anos	2 (13,3)	69 (7,3)	2,1 (0,3 – 10,2)	0,37
Idade				
15-24 anos	0 (0,0)	138 (11,7)	0,0 (-)	0,99
25-40 anos	10 (52,6)	537 (43,7)	(Ref)	(Ref)
41-59 anos	9 (47,4)	507 (41,3)	1,2 (0,4 – 3,7)	0,69
>59 anos	0 (0,0)	45 (3,7)	0,0 (-)	0,99
Ano de Coleta				
2008-2013	6 (31,6)	540 (44,0)	(Ref)	(Ref)
2014-2017	13 (68,4)	687 (56,0)	1,6 (0,5 – 5,7)	0,39
Subtipo				
B	0 (0,0)	212 (17,3)	0,0 (-)	0,99
C	19 (100,0)	898 (73,2)	(Ref)	(Ref)
F1	0 (0,0)	64 (5,2)	0,0 (-)	0,99
CRF31_BC	0 (0,0)	53 (4,3)	0,0 (-)	0,99
Região Intermediária				
Blumenau	5 (26,4)	432 (35,2)	(Ref)	(Ref)
Caçador	1 (5,3)	18 (1,5)	6,3 (0,3 – 50,8)	0,11
Chapecó	2 (10,5)	69 (5,6)	3,3 (0,4 – 18,5)	0,18
Criciúma	3 (15,8)	161 (13,1)	1,8 (0,3 – 8,7)	0,44
Florianópolis	4 (21,0)	305 (24,9)	1,2 (0,2 – 6,5)	0,81
Joinville	2 (10,5)	178 (14,5)	0,9 (0,0 – 6,1)	0,90
Lages	2 (10,5)	64 (5,2)	3,2 (0,4 – 18,8)	0,20

8.6 Análise das redes de transmissão formadas exclusivamente por indivíduos do sexo masculino

Em vista da importância do grupo HSH na dinâmica de transmissão do HIV, *clusters* formados exclusivamente por pacientes do sexo masculino da RENAGENO-SC foram analisados buscando identificar fatores epidemiológicos associados à clusterização desses indivíduos. Dos 1.460 homens amostrados no *dataset*, 183 (12,5%) clusterizaram-se exclusivamente com outros homens, formando 84 *clusters* (25,8% do total de *clusters*). Indivíduos com 12 ou mais anos de escolaridade mostraram uma chance 3,8 vezes maior de estarem em um *cluster* compostos exclusivamente por homens, sendo essa escolaridade encontrada em 19,5% dos homens clusterizados. Quando avaliado os subtipos virais relacionados à infecção desses indivíduos, embora o subtipo C tenha sido mais prevalente tanto nos *clusters* de homens quanto em homens em *clusters* mistos (67,8 e 65,3%, respectivamente), indivíduos infectados com o HIV-1B apresentaram 4,8 vezes mais chances de clusterizarem com outros homens ($p < 0,001$) (Tabela 9). Não foram encontradas significâncias estatísticas envolvendo etnia, escolaridade, ano de coleta e região intermediária que pudessem distinguir o agrupamento desses indivíduos.

Quando analisou-se os *clusters* de homens separadamente para cada subtipo é possível observar a importância do subtipo B nessa via de transmissão. Dos 44 *clusters* compostos só por sequências HIV-1B RENAGENO-SC, 52,3% são *clusters* exclusivamente de homens, enquanto que para o HIV-1C essa proporção é de apenas 21,1% e para o CRF31 é 11,1% (dado não mostrado). *Clusters* exclusivos de homens para HIV-1F1 também possuem uma proporção alta (57,1%); entretanto, o pequeno número de *clusters* RENAGENO-SC para este subtipo ($n=7$) limita as interpretações desses resultados.

Tabela 9: Características epidemiológicas associadas a indivíduos do sexo masculino que clusterizaram exclusivamente com homens.

	Homens clusterizados com homens n= 183 (12,5)	Homens em clusters mistos n= 1277 (87,5)	OR (95% IC)	P value
Etnia				
Não-Caucasiano	19 (12,8)	142 (13,0)	0,5 (0,2 – 1,2)	0,13
Caucasiano	129 (87,2)	945 (82,0)	(Ref)	(Ref)
Anos de Escolaridade				
Nenhum	4 (3,0)	32 (3,3)	0,8 (0,1 – 5,5)	0,8
1-3 anos	10 (7,5)	103 (10,8)	1,5 (0,5 – 5,3)	0,46
4-7 anos	47 (35,4)	417 (43,7)	(Ref)	(Ref)
8-11 anos	46 (34,6)	284 (29,7)	1,2 (0,6 – 2,3)	0,56
12 ou mais anos	26 (19,5)	119 (12,5)	3,8 (1,5 – 10,4)	0,007
Idade				
<14 anos	9 (4,9)	124 (9,7)	0,6 (0,1 – 3,9)	0,58
15-24 anos	15 (8,2)	64 (5,0)	1,4 (0,3 – 6,8)	0,64
25-40 anos	91 (49,8)	441 (34,5)	(Ref)	(Ref)
41-59 anos	63 (34,4)	584 (45,8)	0,8 (0,4 – 1,5)	0,44
>59 anos	5 (2,7)	64 (5,0)	1,1 (0,2 – 4,6)	0,91
Ano de Coleta				
2008-2013	63 (36,6)	647 (50,7)	(Ref)	(Ref)
2014-2017	116 (63,4)	630 (49,3)	1,4 (0,8 – 2,6)	0,24
Região Intermediária				
Blumenau	73 (39,9)	415 (32,5)	(Ref)	(Ref)
Caçador	1 (0,5)	18 (1,4)	0,0 (-)	0,99
Chapecó	9 (4,9)	61 (4,8)	0,5 (0,1 – 1,9)	0,34
Criciúma	30 (16,4)	191 (15,0)	1,0 (0,4 – 2,1)	0,98
Florianópolis	39 (21,3)	373 (29,2)	0,6 (0,2 – 1,4)	0,27
Joinville	24 (13,1)	167 (13,1)	0,7 (0,3 – 1,8)	0,49
Lages	7 (3,8)	52 (4,1)	1,7 (0,2 – 12,0)	0,59
Subtipo				
B	49 (26,8)	316 (24,7)	4,8 (2,1 – 11,3)	<0,001
C	124 (67,8)	833 (65,3)	(Ref)	(Ref)
F1	8 (4,4)	74 (5,8)	2,7 (0,3 – 63,6)	0,42
CRF31_BC	2 (1,0)	54 (4,2)	0,8 (0,0 – 12,6)	0,89
Distância Cluster				
0 km	78 (42,6)	54 (26,5)	1,8 (0,9 – 3,8)	0,08
1-50 km	59 (32,2)	59 (28,9)	1,8 (0,8 – 3,9)	0,12

>50 km	46 (25.2)	91 (44.6)	(Ref)	(Ref)
--------	-----------	-----------	-------	-------

9. DISCUSSÃO

A heterogeneidade dos *datasets* utilizados nesse estudo, somados à inexistência de metodologia consenso para análise de *clusters* de transmissão, trouxe um desafio adicional para identificar e descrever as possíveis redes associadas a disseminação e manutenção da epidemia de HIV-1 no estado de Santa Catarina. Somado a isso, esse é o primeiro estudo conduzido no país que abrange a avaliação das redes de transmissão de quatro subtipos distintos na mesma população: HIV-1B, C, F1 e a forma recombinante CRF31_BC.

Um dos fatores associados à clusterização e capaz de impactar o tamanho e número de *clusters* identificados é o ponto de corte de distância genética máxima. A maioria dos estudos já realizados buscando identificar redes de transmissão do HIV-1 utilizam-se de distâncias genética intracluster de 1 a 9% (Kaye et al., 2008; Hughes et al., 2009; Chalmet et al., 2010; Prosperi et al., 2011; Mehta et al., 2012; Junqueira et al., 2016; Yebra et al., 2018). Visando reduzir o viés de seleção desse parâmetro e evitar a seleção de um valor arbitrário, foi calculado um valor de ponto de corte para cada subtipo baseado no limite superior da taxa evolutiva descrita para cada variante do HIV-1 avaliada (B= 0,022, C=0,024, F1=0,018 e CRF31_BC=0,021), multiplicada por 20, representando 10 anos de evolução em cada indivíduo, visto que é o tempo médio de progressão para o quadro de aids na ausência de tratamento. Além disso, a escolha de uma distância genética mais ou menos restrigente direciona a seleção de *clusters*, onde distâncias menores limitam a identificação a eventos recentes, enquanto distâncias maiores retratam redes mais antigas (Division of HIV/AIDS Prevention & Centers for Disease Control, 2018).

A partir disso foi possível determinar que 943 indivíduos (33,0%) amostrados pela RENAGENO-SC durante 2008-2017 estavam contidos em alguma rede de transmissão. A população clusterizada apresentou diferenças quando à idade, onde indivíduos em faixa etária entre 25 a 40 anos possuíam maiores chances de estarem em uma rede de transmissão. Esse achado pode ser explicado pela elevada taxa de incidência de aids em indivíduos com idade

entre 25 a 39 anos para ambos os sexos, variando de 13,8 casos por 100 mil habitantes em mulheres a 50,9 em homens (Ministério da Saúde, 2019).

O ano de coleta também mostrou-se importante característica para agrupamento. Quando comparadas às sequências coletadas entre 2008 a 2013, amostras coletadas após 2014 apresentaram mais chances de estarem inseridas em um *cluster*. Esse resultado é explicado pela mudança do protocolo de tratamento para HIV/aids preconizado pelo Ministério da Saúde, que a partir de 2014 define que o tratamento deve ser iniciado imediatamente após o diagnóstico. Com isso, espera-se que cada vez mais pacientes necessitem do serviço de genotipagem, aumentando o número de sequências por ano e, conseqüentemente, facilitando amostrar indivíduos epidemiologicamente ligados, visto que a densidade amostral aumentará. A análise de correlação também revelou um coeficiente de correlação quase linear ($R = 0.95$) entre o número de sequências amostradas por ano e o número de sequências agrupadas.

Como já foi investigado por diferentes estudos, a densidade amostral é um fator importante associado ao agrupamento de indivíduos epidemiologicamente relacionados (Novitky et al., 2014). A baixa densidade amostral pode ser o motivo do subtipo F1 ter apresentado a redução significativa de 90%, enquanto HIV-1B e CRF31 tiveram o mesmo padrão, contudo com reduções menores (40 e 50%, respectivamente), indicando que possivelmente há algum fator associado a essas diferenças no agrupamento, podendo envolver a própria rede de transmissão em que esses subtipos circulam. Outro fato que pode estar associado a maior taxa de clusterização observada para o subtipo C é a elevada amostragem de indivíduos do sexo feminino nesse *dataset*. O subtipo C mostrou-se nesse estudo (assim como em trabalhos publicados previamente) mais associado a infecções em mulheres, indicando que possivelmente é a forma viral circulando em maior prevalência em indivíduos heterossexuais. Devido a grande presença de mulheres no *dataset* RENAGENO-SC, a densidade amostral das redes de transmissão do HIV-1C pode ser maior em comparação a outros subtipos avaliados, permitindo assim que os métodos empregados reconstruíssem melhor estas redes. Outro

fator que pode estar influenciando o agrupamento do subtipo C é a relevância epidemiológica desse subtipo na região (Schuelter-Trevisol et al., 2007, Loccateli et al., 2007, Brígido et al., 2007, Rodrigues et al., 2010, Gräf et al., 2011, Gräf et al., 2015), fazendo com que a epidemia de subtipo C seja disseminada. Apesar da variável subtipo viral ter apresentado diferença estatisticamente significativa, impactando no agrupamento, quando os subtipos foram avaliados isoladamente, não foram encontrados achados relevantes que pudessem elucidar um agrupamento diferencial.

Para todos os subtipos o tamanho de *cluster* mais identificado foi envolvendo apenas 2 indivíduos (pares de transmissão). A incidência elevada de *clusters* pequenos pode ser oriunda da própria amostragem das sequências. É possível que esses pares de transmissão façam parte de uma rede maior, entretanto, em vista de impossibilidade de amostrar todos os indivíduos infectados, a reconstrução dos *links* epidemiológicos entre os pares é prejudicada. Esse empecilho é encontrado principalmente em estudos envolvendo grandes epidemias (abarcando países ou continentes), onde há uma baixa densidade amostral de indivíduos, bem como em situações de amostragem de sequências em um longo período (Hué et al., 2004), como é o caso desse estudo.

Outro fator limitante desse estudo é a população diferencial onde os dados moleculares foram obtidos. A avaliação de redes de transmissão em pacientes cronicamente infectados é bastante difícil, principalmente devido ao tempo de divergência genética e evolução intra-hospedeiro a que esses vírus estavam submetidos. O somatório desses fatores evolutivos inviabiliza a reconstrução adequada das relações de transmissão, pois a distância genética entre o vírus sequenciado e o vírus que causou primeiramente a infecção é muito maior, reduzindo a chance dessas sequências clusterizarem (Poon, 2016).

Com a finalidade de avaliar as interações epidemiológicas envolvendo o estado de Santa Catarina, foram incluídas sequências internacionais e de outros estados brasileiros na análise de redes de transmissão. Foi observado

que o subtipo B possuía a maior frequência (53,2%) de *clusters* com sequências externas à RENAGENO-SC, sendo a grande maioria com outros estados do Brasil. Estudo conduzido por Junqueira et al. (2016) demonstrou que o subtipo B forma redes de transmissão que frequentemente extrapolam os limites estaduais, corroborando a ideia de que as relações migratórias entre diferentes estados criam uma potencial rede de difusão de HIV-1 pelo país (Junqueira et al., 2016). No estudo apresentado aqui é possível observar que as redes de transmissão do subtipo B tendem a agrupar sequências de outros estados com mais frequências que os outros subtipos. Entretanto, isto pode ser consequência da abundante quantidade de sequências externas analisadas para este subtipo, decorrência da predominância do subtipo B no Brasil em geral.

A partir dos dados obtidos foi possível observar que os estados do Paraná e Rio Grande do Sul são os de maior relevância epidemiológica quando interconexões entre epidemias foram avaliadas, especialmente para os subtipos B e C. Esses achados estão em consenso com os previamente descritos por Bello et al. (2012a), onde revelou-se uma dinâmica complexa envolvendo as três capitais da região sul: Porto Alegre, Florianópolis e Curitiba. Nesse cenário, Florianópolis atua como a principal ponte ou ponto de conexão entre a epidemia desses dois estados, exercendo a função de receber e disseminar linhagens virais para as cidades vizinhas. Além disso, a migração viral direta descrita entre Porto Alegre e Curitiba é muito baixa, reforçando esse cenário. Outros trabalhos também reforçam a existência de interconexão entre Santa Catarina e o Rio Grande do Sul como, por exemplo, a disseminação de CRF31_BC (Fritsch et al., 2018) e a dispersão de subtipo C, oriundos de Porto Alegre, para outras cidades desse estado (Gräf et al., 2015).

Foram encontradas associações internacionais para o HIV-1C com destaque para Argentina e Alemanha. Esse achado é corroborado por estudo conduzido em países do Cone Sul (Uruguai, Argentina e Paraguai) que indicam a circulação de subtipo C e formas recombinantes BC, principalmente associados a regiões de fronteira com estados da região sul ou por possuir parceiro sexual do Brasil (Carrion et al, 2004, Aulicino et al., 2005a e 2005b,

Montano et al., 2005, Pando et al., 2006, Aguayo et al., 2008, Ruchansky et al., 2009, Pando et al., 2011). A Alemanha também mostrou-se intimamente relacionada com Santa Catarina. Em estudo conduzido por Meixenberger et al (2014) foram relatadas 25 (1,9%) sequências de subtipo C em população recentemente soroconvertida. Embora o estudo não relate com precisão a origem dessas sequências (indivíduos alemães ou imigrantes), 205 (15,4%) dos indivíduos analisados eram de outros países. Dessa forma, não podemos descartar que as sequências alemãs agrupadas com a RENAGENO-SC possam ser de brasileiros imigrantes naquele país, sendo necessário mais estudos para melhor discriminar as possíveis interconexões entre estas epidemias.

Ao avaliar a distribuição e abrangência geoespacial dos *clusters* foi possível observar que a grande maioria dos mesmos era composto por indivíduos morando na mesma cidade, sendo um indicativo de que as redes intramunicipais possuem um papel importante na manutenção da epidemia. Além disso, observa-se que o subtipo C apresentou as maiores distâncias entre as cidades de residências dos membros compondo os *clusters* quando comparado a outros subtipos. Isso pode ser explicado pelo tamanho maior dos *clusters* de subtipo C, que acabam abrangendo indivíduos que moram em cidades diferentes. A reconstrução de *clusters* maiores pode ser consequência de uma melhor representatividade deste subtipo no banco de dados RENAGENO-SC, como já discutido aqui. É ainda importante lembrar que nas análises de distância geoespacial foram considerados apenas *clusters* internos da RENAGENO-SC, pois a informação da cidade de residência dos indivíduos que geraram as sequências externas raramente estava disponível. Desta forma, mais da metade dos *clusters* de subtipo B não foram avaliados nesta análise pois continham sequências externas, onde a localidade do município de origem era incerta.

Além da distribuição geoespacial, alguns aspectos importantes para a manutenção e monitoramento da epidemia de HIV/aids em Santa Catarina foram avaliados, dentre eles a possível transmissão materno-fetal. A presença de infecção por HIV-1 em indivíduos abaixo de 14 anos pode estar associado a

falhas no diagnóstico e acompanhamento pré-natal, além de indicar possível transmissão vertical. No Brasil, no período de 2000 a junho de 2019, foram notificadas mais de 125 mil gestantes infectadas com HIV-1, sendo 30% desses relatos na região Sul, onde a taxa de detecção é o dobro da nacional. Em um período de dez anos, houve um aumento na taxa de detecção dessas mulheres de 2,1 casos/ 1.000 nascidos vivos para 2,9, estando esse dado associado, em parte, pela ampliação de diagnóstico pré-natal da infecção e prevenção da transmissão vertical (Ministério da Saúde, 2019). Embora o número de gestantes infectadas tenha apresentado um pequeno aumento, a incidência de HIV-1 em crianças com idade abaixo de 5 anos vem diminuindo no país, passando de 3,6 em 2008 para 1,9 por mil nascidos vivos em 2018. Apesar dos esforços em reduzir a transmissão vertical, reduzindo a taxa de detecção de 5,3 em 2007 para 2,1 em 2018, o estado de Santa Catarina ainda apresenta taxa de detecção em gestantes três vezes maior que a brasileira e a capital Florianópolis apresenta taxa de detecção de aids em menores de 5 anos de 11,9 casos a cada 100 mil habitantes.

Devido ao fato da genotipagem do HIV-1 ser recomendada para todas as gestantes infectadas, o *dataset* da RENAGENO é uma ótima fonte de dados para se estudar os fatores impactantes na transmissão vertical. Neste estudo foram identificados 19 possíveis pares de mulheres transmissoras e crianças com menos de 14 anos infectadas. Entretanto não foi possível identificar nenhuma característica epidemiológica marcante que pudesse distinguir as possíveis mães transmissoras das mulheres em geral amostradas no RENAGENO-SC. Dentre as crianças agrupadas foi possível identificar que crianças com idade de zero a 2 anos clusterizaram mais, enquanto crianças de 12 a 14 anos clusterizaram menos. Esse dado é um indicativo de que o tempo de evolução intra-hospedeiro do vírus é um forte determinante na capacidade de detectarmos redes de transmissão, ou neste caso, de eventos de transmissão vertical. Em outras palavras, quanto mais próximo ao momento de infecção for feita a coleta e o sequenciamento do vírus, maior é o poder das ferramentas filogenéticas de detectar as redes de transmissão na população estudada.

Além disso, ainda merece uma discussão mais aprofundada o fato de que todos 19 pares de transmissão formados por possíveis mães e crianças estavam infectadas com o HIV-1C. Embora não seja consenso na literatura, estudo conduzido por Odaibo et al. (2006) indicou que a transmissão materno fetal do HIV-1 pode ser influenciada pelo subtipo viral. Além disso, Renjifo e cols (2004) demonstrou que a transmissão intrauterina do HIV-1 ocorre mais frequentemente quando a infecção é associada ao subtipo C. Estudo conduzido por Tornatore et al (2010), buscando elucidar e determinar o momento de transmissão vertical (intraútero, intraparto e pós-parto) em grávidas soropositivas da cidade de Rio Grande (RS), indicou uma taxa de transmissão vertical de aproximadamente 5% (7 casos de 144), sendo 4 casos associados ao subtipo C (3 intraútero e 1 intraparto). É claro que não se pode descartar que a elevada observação de subtipo C em casos de transmissão vertical seja apenas decorrente da elevada frequência deste subtipo entre heterossexuais no sul do Brasil, sendo mais estudos necessários para elucidar tais hipóteses. Por fim, esses dados indicam a necessidade de ampliar o acompanhamento dessas gestantes e garantir acesso e adesão ao tratamento antiretroviral visando reduzir as chances de infecção intraútero e no momento do parto.

Em vista da associação encontrada entre homens e infecção ocasionada por HIV-1B no sul do Brasil, outro dado explorado foi o agrupamento de possíveis indivíduos HSH. A porcentagem de casos em HSH subiu significativamente ao longo dos anos no Brasil, passando de 19,9 em 2007 para 31,7 em 2019 (Ministério da Saúde, 2019). Embora a epidemia nacional seja majoritariamente causada por contato heterossexual, o aumento de casos envolvendo a população HSH é indicativo de vulnerabilidade e o estudo das redes de transmissão dessa população podem ajudar a elucidar fatores de risco para a infecção.

O presente estudo identificou 84 *clusters* compostos exclusivamente de homens, agrupando um total de 183 indivíduos. A associação entre subtipo B e homens manteve-se na análise diferencial de clusterização, onde homens infectados pelo HIV-1B apresentaram 4,8 vezes mais chances de estarem

contidos em um possível *cluster* HSH em comparação a indivíduos infectados pelo subtipo C. Embora o estudo indique forte associação entre subtipo B e uma possível população HSH, o HIV-1C foi mais prevalente, estando presente em 67,8% dos indivíduos em possíveis *clusters* HSH. Esse achado pode estar relacionado à passagem de subtipo C da população heterossexual para a HSH, circulando, atualmente, em elevada prevalência em ambas as populações. Essa hipótese foi investigada por Gräf et al. (2016), onde análises em filodinâmica demonstraram uma elevada taxa de transição de HIV-1C da categoria heterossexual para HSH a partir da década de 1990. Neste estudo também se observa em Santa Catarina uma prevalência de 65,8% de HIV-1C entre HSHs, apesar do HIV-1B estar ainda estatisticamente associado a este grupo. Além disso também foi observado a expansão do HIV-1C entre HSH na cidade de São Paulo (Brígido et al., 2011).

Os dados aqui apresentados favorecem a hipótese que a separação parcial das redes de transmissão é o que molda a segregação diferencial de subtipos do HIV-1 em grupos populacionais, enquanto diferenças biológicas das cepas virais que poderiam favorecer certas vias de transmissão devem ter um impacto menor. Em Santa Catarina, a expansão do HIV-1C parece ocorrer de forma generalizada, em todos os grupos populacionais, enquanto que o HIV-1B parece manter-se forte em redes de transmissão mais restritas e que ainda necessitam serem melhor estudadas. A associação entre elevado grau de escolaridade (12 anos ou mais) e possíveis *clusters* de indivíduos HSH pode sugerir uma seletividade maior dessas redes. Entretanto, o elevado nível educacional dessa população pode também significar uma consciência maior dos riscos de exposição e levar esses indivíduos a procurarem mais rapidamente serviços de saúde e testagem, assim como de genotipagem. Como observado também no agrupamento de crianças de até dois anos, um pequeno tempo entre infecção e amostragem maximiza a capacidade de identificação de redes de transmissão.

Em vista dos resultados apresentados aqui, é possível concluir que a utilização de sequências provenientes do banco de dados da RENAGENO -SC permite identificar e caracterizar redes de transmissão responsáveis por manter

e disseminar a epidemia de HIV-1 pelo estado e outras localidades. Contudo, fica evidente a importância de dados mais detalhados dos pacientes e densidades amostrais mais altas (como as obtidas na RENAGENO-SC a partir de 2014), que possibilitariam detectar e descrever melhor as redes de transmissão.

10. SUMÁRIO DE RESULTADOS

- O HIV-1C apresentou crescimento significativo em prevalência ao longo de dez anos de amostragem.
- A infecção por HIV-1B foi associada a homens, idade entre 25 e 40 anos e elevada escolaridade (12 anos ou mais), enquanto o HIV-1C mostrou-se relacionado a mulheres, indivíduos jovens (idade inferior a 24 anos) e baixa escolaridade (4 a 7 anos).
- As redes de transmissão identificadas neste *dataset* estavam associadas a indivíduos entre 25 e 40 anos, infectados com subtipo C e coletados a partir de 2014.
- HIV-1C apresentou a maior taxa de clusterização, que pode ser devido a alta densidade amostral de mulheres no *dataset* estudado.
- Metade das redes de transmissão do HIV-1B apresentam sequências externas a Santa Catarina, sugerindo uma abrangência nacional destas redes.
- Redes de transmissão intra-estadual são formadas majoritariamente por indivíduos morando na mesma cidade.
- O HIV-1C formou redes maiores e que se estendiam por uma extensão territorial maior dentro do estado de Santa Catarina quando comparado a outros subtipos.
- O HIV-1C mostrou-se predominante em possíveis casos de transmissão vertical identificados.
- Crianças de até dois anos tenderam a clusterizar, mostrando o potencial da abordagem metodológica usada neste estudo para identificar transmissões verticais.
- Redes de transmissão compostas exclusivamente por homens eram associadas a infecção por HIV-1B e alta escolaridade.

11. CONCLUSÃO

O presente estudo analisou 10 anos de amostragem do banco de dados da RENAGENO no estado de Santa Catarina. Este trabalho, é o primeiro a explorar a utilidade deste banco público, porém de acesso restrito, na investigação da epidemiologia molecular e da dinâmica de transmissão do HIV-1. Fica aqui demonstrada que, mesmo em sequências coletadas de indivíduos em infecção crônica, a utilização da abordagem filogenética aliada à aplicação de ferramentas de identificação de *clusters*, pode auxiliar na identificação de redes de transmissão viral e revelar características sobre a circulação do HIV-1 em diferentes subpopulações. Entretanto, ressaltamos que o acesso a dados mais detalhados dos pacientes e a densidades amostrais mais altas (que podem surgir a partir das mudanças nos protocolos de tratamento e notificação da doença) são essenciais para a geração de resultados mais informativos à saúde pública e que finalmente podem ser aplicados ao combate da epidemia do HIV-1.

12. PERSPECTIVAS

- Proceder à análise de dinâmica de transmissão para o único agrupamento com dez indivíduos amostrados, procurando descrever sua taxa de crescimento temporal, além de caracterizar a dispersão dessa rede.
- Avaliar se fatores epidemiológicos dos indivíduos dentro do agrupamento estão associados à dispersão espacial do mesmo.
- Avaliar se a distância genética média dos agrupamentos está associada a composição dos agrupamentos, sua dispersão espacial ou subtipo viral.

REFERÊNCIAS BIBLIOGRÁFICAS

ABECASIS, A. B.; VANDAMME, A. M.; LEMEY, P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. **J Virol**, 83, n. 24, p. 12917-12924, Dec 2009.

ABRAHA, A.; NANKYA, I. L.; GIBSON, R.; DEMERS, K. *et al.* CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic. **J Virol**, 83, n. 11, p. 5592-5605, Jun 2009.

AGUAYO, N.; LAGUNA-TORRES, V. A.; VILLAFANE, M.; BARBOZA, A. *et al.* Epidemiological and molecular characteristics of HIV-1 infection among female commercial sex workers, men who have sex with men and people living with AIDS in Paraguay. **Rev Soc Bras Med Trop**, 41, n. 3, p. 225-231, 2008 May-Jun 2008.

ALFARO, M. E.; ZOLLER, S.; LUTZONI, F. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. **Mol Biol Evol**, 20, n. 2, p. 255-266, Feb 2003.

ALMEIDA, S. E.; DE MEDEIROS, R. M.; JUNQUEIRA, D. M.; GRÄF, T. *et al.* Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. **Virol J**, 9, p. 306, 2012.

ARIËN, K. K.; VANHAM, G.; ARTS, E. J. Is HIV-1 evolving to a less virulent form in humans? **Nat Rev Microbiol**, 5, n. 2, p. 141-151, Feb 2007.

AULICINO, P. C.; KOPKA, J.; MANGANO, A. M.; ROCCO, C. *et al.* Circulation of novel HIV type 1 A, B/C, and F subtypes in Argentina. **AIDS Res Hum Retroviruses**, 21, n. 2, p. 158-164, Feb 2005.

AULICINO, P. C.; KOPKA, J.; ROCCO, C.; MANGANO, A. *et al.* Sequence analysis of a South American HIV type 1 BC recombinant. **AIDS Res Hum Retroviruses**, 21, n. 10, p. 894-896, Oct 2005.

BARIDO-SOTTANI, J.; VAUGHAN, T. G.; STADLER, T. Detection of HIV transmission clusters from phylogenetic trees using a multi-state birth-death model. **J R Soc Interface**, 15, n. 146, 09 2018.

BELLO, G.; AFONSO, J. M.; MORGADO, M. G. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. **Infect Genet Evol**, 12, n. 5, p. 1079-1086, Jul 2012.

BELLO, G.; ZANOTTO, P. M.; IAMARINO, A.; GRÄF, T. *et al.* Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. **PLoS One**, 7, n. 4, p. e35649, 2012.

BENNETT, D. E.; CAMACHO, R. J.; OTELEA, D.; KURITZKES, D. R. *et al.* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. **PLoS One**, 4, n. 3, p. e4724, 2009.

BEZEMER, D.; CORI, A.; RATMANN, O.; VAN SIGHEM, A. *et al.* Dispersion of the HIV-1 Epidemic in Men Who Have Sex with Men in the Netherlands: A Combined Mathematical Model and Phylogenetic Analysis. **PLoS Med**, 12, n. 11, p. e1001898; discussion e1001898, Nov 2015.

BEZEMER, D.; VAN SIGHEM, A.; LUKASHOV, V. V.; VAN DER HOEK, L. *et al.* Transmission networks of HIV-1 among men having sex with men in the Netherlands. **AIDS**, 24, n. 2, p. 271-282, Jan 2010.

BRENNER, B. G. *et al.* High rates of forward transmission events after acute/early HIV-1 infection. **J Infect Dis**, v. 195, n. 7, p. 951-9, Apr 2007. ISSN 0022-1899.

BRÍGIDO, L. F. *et al.* HIV type 1 subtype C and CB Pol recombinants prevail at the cities with the highest AIDS prevalence rate in Brazil. **AIDS Res Hum Retroviruses**, v. 23, n. 12, p. 1579-86, Dec 2007. ISSN 0889-2229.

BRÍGIDO, L. F.; FERREIRA, J. L.; ALMEIDA, V. C.; ROCHA, S. Q. *et al.* Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. **AIDS Res Hum Retroviruses**, 27, n. 3, p. 339-344, Mar 2011.

BRINDEIRO, R. M. *et al.* Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. **AIDS**, v. 17, n. 7, p. 1063-9, May 2003. ISSN 0269-9370.

BRITO, A. M.; CASTILHO, E. A.; SZWARCOWALD, C. L. AIDS and HIV infection in Brazil: a multifaceted epidemic. **Rev Soc Bras Med Trop**, 34, n. 2, p. 207-217, 2001 Mar-Apr 2001.

BRODER, S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. **Antiviral Res**, 85, n. 1, p. 1-18, Jan 2010.

CARDOSO, L. P.; QUEIROZ, B. B.; STEFANI, M. M. HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naïve patients from Central West Brazil. **J Clin Virol**, 46, n. 2, p. 134-139, Oct 2009.

CARRION, G.; EYZAGUIRRE, L.; MONTANO, S. M.; LAGUNA-TORRES, V. *et al.* Documentation of subtype C HIV Type 1 strains in Argentina, Paraguay, and Uruguay. **AIDS Res Hum Retroviruses**, 20, n. 9, p. 1022-1025, Sep 2004.

CERQUEIRA, D. M.; AMORIM, R. M.; SILVA, R. R.; CAMARA, G. N. *et al.* Antiretroviral resistance and genetic diversity of human immunodeficiency virus type 1 isolates from the Federal District, Central Brazil. **Mem Inst Oswaldo Cruz**, 99, n. 8, p. 877-882, Dec 2004.

CHALMET, K.; STAELENS, D.; BLOT, S.; DINAKIS, S. *et al.* Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. **BMC Infect Dis**, 10, p. 262, Sep 2010.

COUTO-FERNANDEZ, J. C. *et al.* Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil. **AIDS Res Hum Retroviruses**, v. 22, n. 2, p. 207-11, Feb 2006. ISSN 0889-2229.

CUEVAS, M.; FERNANDEZ-GARCIA, A.; SANCHEZ-GARCIA, A.; GONZALEZ-GALEANO, M. *et al.* Incidence of non-B subtypes of HIV-1 in Galicia, Spain: high frequency and diversity of HIV-1 among men who have sex with men. **Euro Surveill**, 14, n. 47, Nov 2009.

CULLEN, B. R. HIV-1 auxiliary proteins: making connections in a dying cell. **Cell**, 93, n. 5, p. 685-692, May 1998.

D'ORSI, E. **Informativo Epidemiológico "Barriga Verde"**. 2004. *apud* LOCATELI, D. *et al.* Molecular epidemiology of HIV-1 in Santa Catarina State confirms increases of subtype C in Southern Brazil. **J Med Virol**, v. 79, n. 10, p. 1455-63, Oct 2007. ISSN 0146-6615.

DE MATOS, M. A.; CAETANO, K. A.; FRANÇA, D. D.; PINHEIRO, R. S. *et al.* Vulnerability to sexually transmitted infections in women who sell sex on the route of prostitution and sex tourism in central Brazil. **Rev Lat Am Enfermagem**, 21, n. 4, p. 906-912, 2013 Jul-Aug 2013.

DE MEDEIROS, R. M.; JUNQUEIRA, D. M.; MATTE, M. C.; BARCELLOS, N. T. *et al.* Co-circulation HIV-1 subtypes B, C, and CRF31_BC in a drug-naïve population from Southernmost Brazil: analysis of primary resistance mutations. **J Med Virol**, 83, n. 10, p. 1682-1688, Oct 2011.

DENNIS, A. M.; HERBECK, J. T.; BROWN, A. L.; KELLAM, P. *et al.* Phylogenetic studies of transmission dynamics in generalized HIV epidemics: an essential tool where the burden is greatest? **J Acquir Immune Defic Syndr**, 67, n. 2, p. 181-195, Oct 2014.

DE OLIVEIRA, T. *et al.* Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. **Lancet HIV**, v. 4, n. 1, p. e41-e50, 01 2017. ISSN 2352-3018.

DEARLOVE, B. L.; XIANG, F.; FROST, S. D. W. Biased phylodynamic inferences from analysing clusters of viral sequences. **Virus Evol**, 3, n. 2, p. vex020, Jul 2017.

DEBRY, R. W.; ABELE, L. G.; WEISS, S. H.; HILL, M. D. *et al.* Dental HIV transmission? **Nature**, 361, n. 6414, p. 691, Feb 1993.

DELATORRE, E.; COUTO-FERNANDEZ, J. C.; GUIMARÃES, M. L.; VAZ CARDOSO, L. P. *et al.* Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. **PLoS One**, 8, n. 9, p. e74072, 2013.

DIAS, C. F.; NUNES, C. C.; FREITAS, I. O.; LAMEGO, I. S. *et al.* High prevalence and association of HIV-1 non-B subtype with specific sexual transmission risk among antiretroviral naïve patients in Porto Alegre, RS, Brazil. **Rev Inst Med Trop Sao Paulo**, 51, n. 4, p. 191-196, 2009 Jul-Aug 2009.

DIAS, C. F.; NUNES, C. C.; FREITAS, I. O.; LAMEGO, I. S. *et al.* High prevalence and association of HIV-1 non-B subtype with specific sexual transmission risk among antiretroviral naïve patients in Porto Alegre, RS, Brazil. **Rev Inst Med Trop Sao Paulo**, 51, n. 4, p. 191-196, 2009 Jul-Aug 2009.

DIVISION OF HIV/AIDS PREVENTION & CENTERS FOR DISEASE CONTROL. Detecting, investigating and responding to HIV transmission clusters. Centers for Disease Control and Prevention, 2018. Disponível em <<http://test.datamonkey.org/hivtrace>>

EFRON, B.; HALLORAN, E.; HOLMES, S. Bootstrap confidence levels for phylogenetic trees. **Proc Natl Acad Sci U S A**, 93, n. 23, p. 13429-13434, Nov 1996.

ERIXON, P.; SVENNBLAD, B.; BRITTON, T.; OXELMAN, B. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. **Syst Biol**, 52, n. 5, p. 665-673, Oct 2003.

FELSENSTEIN, J. Cases in which parsimony or compatibility methods will be positively misleading. **Syst Zool**, 27:401-4010, 1978.

FELSENSTEIN, J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. **Evolution**, 39, n. 4, p. 783-791, Jul 1985.

FERREIRA, J. L.; THOMAZ, M.; RODRIGUES, R.; HARRAD, D. *et al.* Molecular characterisation of newly identified HIV-1 infections in Curitiba, Brazil: preponderance of clade C among males with recent infections. **Mem Inst Oswaldo Cruz**, 103, n. 8, p. 800-808, Dec 2008.

FITZJOHN, RG. Diversitree: comparative phylogenetic analyses of diversification in R. **Methods Ecol Evol**, 3, 1084-1092, 2012.

FONSECA, M. G.; BASTOS, F. I. Twenty-five years of the AIDS epidemic in Brazil: principal epidemiological findings, 1980-2005. **Cad Saude Publica**, 23 Suppl 3, p. S333-344, 2007.

FOURMENT, M.; GIBBS, M. J. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. **BMC Evol Biol**, 6, p. 1, Jan 2006.

FRITSCH, H. M. et al. Spatiotemporal and demographic history of the HIV-1 circulating recombinant form CRF31_BC in Brazil. **Infect Genet Evol**, v. 61, p. 113-118, Jul 2018.

GADELHA, S. R. et al. Molecular epidemiology of human immunodeficiency virus-1 in the state of Ceará, Northeast, Brazil. **Mem Inst Oswaldo Cruz**, v. 98, n. 4, p. 461-3, Jun 2003. ISSN 0074-0276.

GRABOWSKI, M. K.; REDD, A. D. Molecular tools for studying HIV transmission in sexual networks. **Curr Opin HIV AIDS**, v. 9, n. 2, p. 126-33, Mar 2014. ISSN 1746-6318.

GRÄF, T. et al. Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. **J Virol**, v. 89, n. 24, p. 12341-8, Dec 2015. ISSN 1098-5514.

GRÄF, T. et al. HIV-1 genetic diversity and drug resistance among treatment naïve patients from Southern Brazil: an association of HIV-1 subtypes with exposure categories. **J Clin Virol**, v. 51, n. 3, p. 186-91, Jul 2011. ISSN 1873-5967.

GRÄF, T. et al. Comprehensive characterization of HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS. **J Virol**, v. 90, n. 18, p. 8160-8, 09 2016. ISSN 1098-5514.

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

GRENFELL, B. T. et al. Unifying the epidemiological and evolutionary dynamics of pathogens. **Science**, v. 303, n. 5656, p. 327-32, Jan 2004. ISSN 1095-9203.

GUIMARÃES, M. L. et al. High frequency of recombinant genomes in HIV type 1 samples from Brazilian southeastern and southern regions. **AIDS Res Hum Retroviruses**, v. 18, n. 17, p. 1261-9, Nov 2002. ISSN 0889-2229. .

HARRIS, M. E.; MAAYAN, S.; KIM, B.; ZEIRA, M. et al. A cluster of HIV type 1 subtype C sequences from Ethiopia, observed in full genome analysis, is not sustained in subgenomic regions. **AIDS Res Hum Retroviruses**, 19, n. 12, p. 1125-1133, Dec 2003.

HASSAN, A. S. et al. Defining HIV-1 transmission clusters based on sequence data. **AIDS**, v. 31, n. 9, p. 1211-1222, 06 2017. ISSN 1473-5571.

HEMELAAR, J. et al. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. **AIDS**, v. 20, n. 16, p. W13-23, Oct 2006. ISSN 0269-9370.

HOFFMANN, C.; ROCKSTROH, J.; KAMPS, B. **HIV Medicine**. 15 ed. Paris: Flying Publisher, 2007.

HUÉ, S. et al. Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naïve individuals. **J Virol**, v. 83, n. 6, p. 2645-54, Mar 2009. ISSN 1098-5514.

HUÉ, S.; CLEWLEY, J. P.; CANE, P. A.; PILLAY, D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. **AIDS**, 18, n. 5, p. 719-728, Mar 2004.

HUGHES, G. J. et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. **PLoS Pathog**, v. 5, n. 9, p. e1000590, Sep 2009. ISSN 1553-7374.

IBGE- Instituto Brasileiro de Geografia e Estatística. Cidades: Panorama de estados – Santa Catarina, Brasil, 2017. Disponível em <<https://cidades.ibge.gov.br/brasil/sc/panorama>> . Acesso em jul 2018.

IODANSKIY, S.; WALTKE, M.; FENG, Y.; WOOD, C. Subtype-associated differences in HIV-1 reverse transcription affect the viral replication. **Retrovirology**, 7, p. 85, Oct 2010.

JEFFROY, O.; BRINKMANN, H.; DELSUC, F.; PHILIPPE, H. Phylogenomics: the beginning of incongruence? **Trends Genet**, 22, n. 4, p. 225-231, Apr 2006.

JUNQUEIRA, D. M. et al. Short-Term dynamic and local Epidemiological Trends in the South American HIV-1B Epidemic. **PLoS One**, v. 11, n. 6, p. e0156712, 2016. ISSN 1932-6203.

JUNQUEIRA, D. M.; SIBISI, Z.; WILKINSON, E.; DE OLIVEIRA, T. Factors influencing HIV-1 phylogenetic clustering. **Curr Opin HIV AIDS**, 14, n. 3, p. 161-172, May 2019.

KALYAANAMOORTHY, S.; MINH, B. Q.; WONG, T. K. F.; VON HAESELER, A. et al. ModelFinder: fast model selection for accurate phylogenetic estimates. **Nat Methods**, 14, n. 6, p. 587-589, Jun 2017.

KANTOR, R.; SMEATON, L.; VARDHANABHUTI, S.; HUDELSON, S. E. et al. Pretreatment HIV Drug Resistance and HIV-1 Subtype C Are Independently Associated With Virologic Failure: Results From the Multinational PEARLS (ACTG A5175) Clinical Trial. **Clin Infect Dis**, 60, n. 10, p. 1541-1549, May 2015.

KATOH, K.; MISAWA, K.; KUMA, K.; MIYATA, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Res**, 30, n. 14, p. 3059-3066, Jul 2002.

KAYE, M.; CHIBO, D.; BIRCH, C. Phylogenetic investigation of transmission pathways of drug-resistant HIV-1 utilizing pol sequences derived from resistance genotyping. **J Acquir Immune Defic Syndr**, 49, n. 1, p. 9-16, Sep 2008.

KOSAKOVSKY POND, S. L.; WEAVER, S.; LEIGH BROWN, A. J.; WERTHEIM, J. O. HIV-TRACE (TRANSMISSION Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. **Mol Biol Evol**, 35, n. 7, p. 1812-1819, 07 2018.

KÜHNERT, D.; STADLER, T.; VAUGHAN, T. G.; DRUMMOND, A. J. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. **Mol Biol Evol**, 33, n. 8, p. 2102-2116, 08 2016.

LARSSON, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics**, v. 30, n. 22, p. 3276-8, Nov 2014. ISSN 1367-4811.

LAU, K. A.; WONG, J. J. Current Trends of HIV Recombination Worldwide. **Infect Dis Rep**, 5, n. Suppl 1, p. e4, Jun 2013.

LAZARINI, F. M.; MELCHIOR, R.; GONZÁLEZ, A. D.; MATSUO, T. Trends in the epidemic of Aids cases in Southern Brazil from 1986 to 2008. **Rev Saude Publica**, 46, n. 6, p. 960-968, Dec 2012.

LEMEY, P.; VANDAMME, A. M. Exploring full-genome sequences for phylogenetic support of HIV-1 transmission events. **AIDS**, 19, n. 14, p. 1551-1552, Sep 2005.

LEMOINE, F.; DOMELEVO ENTFELLNER, J. B.; WILKINSON, E.; CORREIA, D. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. **Nature**, 556, n. 7702, p. 452-456, 04 2018.

LEWIS, F. *et al.* Episodic sexual transmission of HIV revealed by molecular phylodynamics. **PLoS Med**, v. 5, n. 3, p. e50, Mar 2008. ISSN 1549-1676.

LOCATELI, D. *et al.* Molecular epidemiology of HIV-1 in Santa Catarina State confirms increases of subtype C in Southern Brazil. **J Med Virol**, v. 79, n. 10, p. 1455-63, Oct 2007. ISSN 0146-6615.

LOLE, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. **J Virol**, v. 73, n. 1, p. 152-60, Jan 1999. ISSN 0022-538X.

LOS ALAMOS NATIONAL LABORATORY. Los Alamos HIV Sequence Database. **Los Alamos National Laboratory, NM.**, 2018. Disponível em: <<https://www.hiv.lanl.gov/>>. Acesso em: jun. de 2018.

LUBELCHEK, R. J. et al. Transmission clustering among newly diagnosed HIV patients in Chicago, 2008 to 2011: using phylogenetics to expand knowledge of regional HIV transmission patterns. **J Acquir Immune Defic Syndr**, v. 68, n. 1, p. 46-54, Jan 2015. ISSN 1944-7884.

MACHADO, L. F. et al. Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. **AIDS Res Hum Retroviruses**, v. 25, n. 10, p. 961-6, Oct 2009. ISSN 1931-8405.

MADDISON, W. P.; MIDFORD, P. E.; OTTO, S. P. Estimating a binary character's effect on speciation and extinction. **Syst Biol**, 56, n. 5, p. 701-710, Oct 2007.

MARZEL, A.; SHILAIH, M.; YANG, W. L.; BÖNI, J. *et al.* HIV-1 Transmission During Recent Infection and During Treatment Interruptions as Major Drivers of New Infections in the Swiss HIV Cohort Study. **Clin Infect Dis**, 62, n. 1, p. 115-122, Jan 2016.

MEHTA, S. R.; DELPORT, W.; BROUWER, K. C.; ESPITIA, S. *et al.* The relatedness of HIV epidemics in the United States-Mexico border region. **AIDS Res Hum Retroviruses**, 26, n. 12, p. 1273-1277, Dec 2010.

MEHTA, S. R.; KOSAKOVSKY POND, S. L.; YOUNG, J. A.; RICHMAN, D. *et al.* Associations between phylogenetic clustering and HLA profile among HIV-infected individuals in San Diego, California. **J Infect Dis**, 205, n. 10, p. 1529-1533, May 2012.

MEIXENBERGER, K.; HAUSER, A.; JANSEN, K.; YOUSEF, K. P. *et al.* Assessment of ambiguous base calls in HIV-1 pol population sequences as a biomarker for identification of recent infections in HIV-1 incidence studies. **J Clin Microbiol**, 52, n. 8, p. 2977-2983, Aug 2014.

MELLORS, J. W. et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. **Ann Intern Med**, v. 122, n. 8, p. 573-9, Apr 1995. ISSN 0003-4819.

MILLER, M.A. et al. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. **Proceedings of the Gateway Computing Environments Workshop (GCE)**. New Orleans; 2010, pp, 1-8.

MILLER, R. H.; SARVER, N. HIV accessory proteins as therapeutic targets. **Nat Med**, v. 3, n. 4, p. 389-94, Apr 1997. ISSN 1078-8956.

MINISTÉRIO DA SAÚDE; Coordenação de DST e Aids. Brasil. **Boletim Epidemiológico AIDS**. Ano V, no. 1, até semana epidemiológica 26a – Junho de 2017. Brasília; 2017.

MINISTÉRIO DA SAÚDE; Coordenação de DST e Aids. Brasil. **Protocolo Clínico e Diretrizes Terapêuticas para Manejo da Infecção pelo HIV em Adultos**. 1 ed. Brasília; 2018.

MINISTÉRIO DA SAÚDE; Coordenação de DST e Aids. Brasil. **Protocolo Clínico e Diretrizes Terapêuticas para Manejo da Infecção pelo HIV em Adultos**. 1 ed. Brasília; 2019.

MINISTÉRIO DA SAÚDE; Sistema de Informação de Agravos de Notificação. Brasil. **Dados Epidemiológicos Nacionais, Regionais e Municipais**. Brasília, 2017.

MINISTÉRIO DA SAÚDE; Sistema de Informação de Agravos de Notificação. Brasil. **Dados Epidemiológicos Nacionais, Regionais e Municipais**. Brasília, 2019.

MIR, D.; CABELLO, M.; ROMERO, H.; BELLO, G. Phylodynamics of major HIV-1 subtype B pandemic clades circulating in Latin America. **AIDS**, 29, n. 14, p. 1863-1869, Sep 2015.

MONTANO, S. M.; SANCHEZ, J. L.; LAGUNA-TORRES, A.; CUCHI, P. *et al.* Prevalences, genotypes, and risk factors for HIV transmission in South America. **J Acquir Immune Defic Syndr**, 40, n. 1, p. 57-64, Sep 2005.

MONTEIRO, J. P.; FERRARO, G. A.; OLIVEIRA, T.; GOLDANI, L. Z. *et al.* Genetic and biologic characterization of HIV type 1 subtype C isolates from south Brazil. **AIDS Res Hum Retroviruses**, 23, n. 1, p. 135-143, Jan 2007.

NAIDOO, L.; MZOBE, Z.; JIN, S. W.; RAJKOOMAR, E. *et al.* Nef-mediated inhibition of NFAT following TCR stimulation differs between HIV-1 subtypes. **Virology**, 531, p. 192-202, 05 2019.

NGUYEN, L. T.; SCHMIDT, H. A.; VON HAESLER, A.; MINH, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Mol Biol Evol**, 32, n. 1, p. 268-274, Jan 2015.

NOVITSKY, V.; MOYO, S.; LEI, Q.; DEGRUTTOLA, V. *et al.* Impact of sampling density on the extent of HIV clustering. **AIDS Res Hum Retroviruses**, 30, n. 12, p. 1226-1235, Dec 2014.

NOVITSKY, V.; MOYO, S.; LEI, Q.; DEGRUTTOLA, V. *et al.* Importance of Viral Sequence Length and Number of Variable and Informative Sites in Analysis of HIV Clustering. **AIDS Res Hum Retroviruses**, 31, n. 5, p. 531-542, May 2015.

NUNES, C. C.; MATTE, M. C.; DIAS, C. F.; ARAÚJO, L. A. *et al.* The influence of HIV-1 subtypes C, CRF31_BC and B on disease progression and initial virologic response to HAART in a Southern Brazilian cohort. **Rev Inst Med Trop Sao Paulo**, 56, n. 3, p. 205-211, 2014 May-Jun 2014.

ODAIBO, G. N.; OLALEYE, D. O.; HEYNDRICKX, L.; VEREECKEN, K. *et al.* Mother-to-child transmission of different HIV-1 subtypes among ARV Naïve infected pregnant women in Nigeria. **Rev Inst Med Trop Sao Paulo**, 48, n. 2, p. 77-80, 2006 Mar-Apr 2006.

OSMANOV, S. *et al.* Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. **J Acquir Immune Defic Syndr**, v. 29, n. 2, p. 184-90, Feb 2002. ISSN 1525-4135.

PANDO, M. A.; EYZAGUIRRE, L. M.; SEGURA, M.; BAUTISTA, C. T. *et al.* First report of an HIV-1 triple recombinant of subtypes B, C and F in Buenos Aires, Argentina. **Retrovirology**, 3, p. 59, Sep 2006.

PANDO, M. A.; GÓMEZ-CARRILLO, M.; VIGNOLES, M.; RUBIO, A. E. *et al.* Incidence of HIV type 1 infection, antiretroviral drug resistance, and molecular characterization in newly diagnosed individuals in Argentina: A Global Fund Project. **AIDS Res Hum Retroviruses**, 27, n. 1, p. 17-23, Jan 2011.

PASSAES, C. P.; BELLO, G.; LORETE, R. S.; MATOS ALMEIDA, S. E. *et al.* Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in Southern Brazil. **Infect Genet Evol**, 9, n. 4, p. 474-482, Jul 2009.

PEDROSO, C. *et al.* High prevalence of primary antiretroviral resistance among HIV-1-infected adults and children in Bahia, a northeast state of Brazil. **J Acquir Immune Defic Syndr**, v. 45, n. 2, p. 251-3, Jun 2007.

PEÑA, A. C. P. *et al.* Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of new REGA version 3 and seven other tools. **Infect Genet Evol**, p. 337-348, doi: 10.1016/j.meegid.2013.04.032.

PEREIRA, G. A. *et al.* Human immunodeficiency virus type 1 and hepatitis C virus co-infection and viral subtypes at an HIV testing center in Brazil. **J Med Virol**, v. 78, n. 6, p. 719-23, Jun 2006. ISSN 0146-6615.

PERELSON, A. S. *et al.* HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. **Science**, v. 271, n. 5255, p. 1582-6, Mar 1996. ISSN 0036-8075.

PILON, R.; LEONARD, L.; KIM, J.; VALLEE, D. *et al.* Transmission patterns of HIV and hepatitis C virus among networks of people who inject drugs. **PLoS One**, 6, n. 7, p. e22245, 2011.

PLANTIER, J. C. et al. A new human immunodeficiency virus derived from gorillas. **Nat Med**, v. 15, n. 8, p. 871-2, Aug 2009. ISSN 1546-170X.

POON, A. F. et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. **Lancet HIV**, v. 3, n. 5, p. e231-8, 05 2016. ISSN 2352-3018.

PRESTON, B. D.; POIESZ, B. J.; LOEB, L. A. Fidelity of HIV-1 reverse transcriptase. **Science**, 242, n. 4882, p. 1168-1171, Nov 1988.

PROSPERI, M. C. et al. A novel methodology for large-scale phylogeny partition. **Nat Commun**, v. 2, p. 321, 2011. ISSN 2041-1723.

RABONI, S. M. et al. Molecular epidemiology of HIV-1 clades in Southern Brazil. **Mem Inst Oswaldo Cruz**, v. 105, n. 8, p. 1044-9, Dec 2010a. ISSN 1678-8060.

RAGONNET-CRONIN, M. et al. Automated analysis of phylogenetic clusters. **BMC Bioinformatics**, v. 14, p. 317, Nov 2013. ISSN 1471-2105.

RAMBAUT, A. et al. The causes and consequences of HIV evolution. **Nat Rev Genet**, v. 5, n. 1, p. 52-61, Jan 2004. ISSN 1471-0056.

RAMOS, A. et al. Dual and recombinant infections: an integral part of the HIV-1 epidemic in Brazil. **Emerg Infect Dis**, v. 5, n. 1, p. 65-74, 1999 Jan-Feb 1999. ISSN 1080-6040.

RATMANN, O. et al. Sources of HIV infection among men having sex with men and implications for prevention. **Sci Transl Med**, v. 8, n. 320, p. 320ra2, Jan 2016. ISSN 1946-6242.

RENJIFO, B.; GILBERT, P.; CHAPLIN, B.; MSAMANGA, G. et al. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. **AIDS**, 18, n. 12, p. 1629-1636, Aug 2004.

RODRIGUES, R. et al. Low prevalence of primary antiretroviral resistance mutations and predominance of HIV-1 clade C at polymerase gene in newly diagnosed individuals from south Brazil. **Virus Res**, v. 116, n. 1-2, p. 201-7, Mar 2006. ISSN 0168-1702.

RODRIGUES, R.; MANENTI, S.; ROMAO, P. R.; DE PAULA FERREIRA, J. L. et al. Young pregnant women living with HIV/AIDS in Criciuma, Southern Brazil, are infected almost exclusively with HIV type 1 clade C. **AIDS Res Hum Retroviruses**, 26, n. 3, p. 351-357, Mar 2010.

ROKAS, A.; CHATZIMANOLIS, S. From gene-scale to genome-scale phylogenetics: the data flood in, but the challenges remain. **Methods Mol Biol**, 422, p. 1-12, 2008.

ROKAS, A.; WILLIAMS, B. L.; KING, N.; CARROLL, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. **Nature**, 425, n. 6960, p. 798-804, Oct 2003.

RUCHANSKY, D.; CASADO, C.; RUSSI, J. C.; ARBIZA, J. R. *et al.* Identification of a new HIV Type 1 circulating recombinant form (CRF38_BF1) in Uruguay. **AIDS Res Hum Retroviruses**, 25, n. 3, p. 351-356, Mar 2009.

SA FILHO, D. J. *et al.* Analysis of full-length human immunodeficiency virus type 1 genome reveals a variable spectrum of subtypes B and f recombinants in São Paulo, Brazil. **AIDS Res Hum Retroviruses**, v. 21, n. 2, p. 145-51, Feb 2005. ISSN 0889-2229.

SANTOS, A. F. *et al.* Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. **AIDS**, v. 20, n. 16, p. 2011-9, Oct 2006. ISSN 0269-9370.

SCHUELTER-TREVISOL, F.; DA SILVA, M. V.; OLIVEIRA, C. M.; RODRIGUES, R. HIV genotyping among female sex workers in the State of Santa Catarina. **Rev Soc Bras Med Trop**, 40, n. 3, p. 259-263, 2007 May-Jun 2007.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2003. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2004. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2005. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2006. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2007. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário

Estatístico de Turismo de Santa Catarina. Ano 2008. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2009. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2010. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2011. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2012. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2013. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2014. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2015. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2016. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SECRETARIA ESTADUAL DE TURISMO DE SANTA CATARINA. Anuário Estatístico de Turismo de Santa Catarina. Ano 2017. Disponível em <<http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>>. Acesso em Março/2019.

SILVA, M. M. et al. HIV subtype, epidemiological and mutational correlations in patients from Paraná, Brazil. **Braz J Infect Dis**, v. 14, n. 5, p. 495-501, 2010 Sep-Oct 2010.

SILVEIRA, J. et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. **J Clin Virol**, v. 54, n. 1, p. 36-41, May 2012. ISSN 1873-5967.

SIMON, D. et al. Prevalence of HIV-1 subtypes in patients of an urban center in Southern Brazil. **Rev Saude Publica**, v. 44, n. 6, p. 1094-101, Dec 2010. ISSN 1518-8787.

SOARES, E. A. et al. Epidemiologic and molecular characterization of human immunodeficiency virus type 1 in southern Brazil. **J Acquir Immune Defic Syndr**, v. 34, n. 5, p. 520-6, Dec 2003. ISSN 1525-4135.

SOARES, E. A.; MARTÍNEZ, A. M.; SOUZA, T. M.; SANTOS, A. F. et al. HIV-1 subtype C dissemination in southern Brazil. **AIDS**, 19 Suppl 4, p. S81-86, Oct 2005.

SOARES, M. A.; DE OLIVEIRA, T.; BRINDEIRO, R. M.; DIAZ, R. S. et al. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. **AIDS**, 17, n. 1, p. 11-21, Jan 2003.

STADLER, T.; BONHOEFFER, S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. **Philos Trans R Soc Lond B Biol Sci**, 368, n. 1614, p. 20120198, Mar 2013.

STRUCK, D. et al. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. **Nucleic Acids Res**, v. 42, n. 18, p. e144, Oct 2014. ISSN 1362-4962.

STRUNNIKOVA, N.; RAY, S. C.; LIVINGSTON, R. A.; RUBALCABA, E. et al. Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. **J Virol**, 69, n. 12, p. 7548-7558, Dec 1995.

STÜRMER, M.; PREISER, W.; GUTE, P.; NISIUS, G. et al. Response to 'Single phylogenetic reconstruction method is insufficient to clarify relationships between patient isolates in HIV-1 transmission case' by Jenwitheesuk and Liu. **AIDS**, 19, n. 7, p. 741-743; author reply 743-744, Apr 2005.

TEBIT, D. M.; ARTS, E. J. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. **Lancet Infect Dis**, 11, n. 1, p. 45-56, Jan 2011.

TEIXEIRA, S. L. et al. HIV-1 infection among injection and ex-injection drug users from Rio de Janeiro, Brazil: prevalence, estimated incidence and genetic

diversity. **J Clin Virol**, v. 31, n. 3, p. 221-6, Nov 2004. ISSN 1386-6532.

TOLEDO, P. V. et al. Genetic diversity of human immunodeficiency virus-1 isolates in Paraná, Brazil. **Braz J Infect Dis**, v. 14, n. 3, p. 230-6, 2010 May-Jun 2010. ISSN 1678-4391.

TORNATORE, M.; GONÇALVES, C. V.; MENDOZA-SASSI, R. A.; SILVEIRA, J. M. et al. HIV-1 vertical transmission in Rio Grande, Southern Brazil. **Int J STD AIDS**, 21, n. 5, p. 351-355, May 2010.

TOVANABUTRA, S.; BEYRER, C.; SAKKHACHORNPHOP, S.; RAZAK, M. H. et al. The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. **AIDS Res Hum Retroviruses**, 20, n. 5, p. 465-475, May 2004.

VASAN, A.; RENJIFO, B.; HERTZMARK, E.; CHAPLIN, B. et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. **Clin Infect Dis**, 42, n. 6, p. 843-852, Mar 2006.

VOLZ, E. M.; FROST, S. D. Inferring the source of transmission with phylogenetic data. **PLoS Comput Biol**, v. 9, n. 12, p. e1003397, 2013. ISSN 1553-7358.

VOLZ, E. M.; IONIDES, E.; ROMERO-SEVERSON, E. O.; BRANDT, M. G. et al. HIV-1 transmission during early infection in men who have sex with men: a phylodynamic analysis. **PLoS Med**, 10, n. 12, p. e1001568; discussion e1001568, Dec 2013.

VOLZ, E. M.; KOOPMAN, J. S.; WARD, M. J.; BROWN, A. L. et al. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. **PLoS Comput Biol**, 8, n. 6, p. e1002552, 2012.

VRBIK, I.; STEPHENS, D. A.; ROGER, M.; BRENNER, B. G. The Gap Procedure: for the identification of phylogenetic clusters in HIV-1 sequence data. **BMC Bioinformatics**, 16, p. 355, Nov 2015.

WANG, Z.; ZHANG, M.; ZHANG, R.; LIU, L. et al. Diversity of HIV-1 genotypes and high prevalence of pretreatment drug resistance in newly diagnosed HIV-infected patients in Shanghai, China. **BMC Infect Dis**, 19, n. 1, p. 313, Apr 2019.

WERTHEIM, J. O.; KOSAKOVSKY POND, S. L.; FORGIONE, L. A.; MEHTA, S. R. et al. Social and Genetic Networks of HIV-1 Transmission in New York City. **PLoS Pathog**, 13, n. 1, p. e1006000, 01 2017.

WERTHEIM, J. O.; SCHEFFLER, K.; CHOI, J. Y.; SMITH, D. M. et al. Phylogenetic relatedness of HIV-1 donor and recipient populations. **J Infect Dis**, 207, n. 7, p. 1181-1182, Apr 2013.

WHO/UNAIDS. **Global report: UNAIDS report on the global AIDS epidemic:** Joint United Nations Programme on HIV/AIDS 2019.

WILBE, K. Genetic dynamics of HIV-1: recombination, drug resistance and intrahost evolution, 2004.

YEBRA, G.; FRAMPTON, D.; GALLO CASSARINO, T.; RAFFLE, J. *et al.* A high HIV-1 strain variability in London, UK, revealed by full-genome analysis: Results from the ICONIC project. **PLoS One**, 13, n. 2, p. e0192081, 2018.

YEBRA, G.; HODCROFT, E. B.; RAGONNET-CRONIN, M. L.; PILLAY, D. *et al.* Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. **Sci Rep**, 6, p. 39489, 12 2016.

YERLY, S. *et al.* The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. **AIDS**, v. 23, n. 11, p. 1415-23, Jul 2009. ISSN 1473-5571.

ZHANG, L. Q.; MACKENZIE, P.; CLELAND, A.; HOLMES, E. C. *et al.* Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. **J Virol**, 67, n. 6, p. 3345-3356, Jun 1993.

APÊNDICES

Suplementar 1

Tabela 10: Comparação de tamanho de clados obtidos com Cluster Picker, a partir de árvore com e sem TBE.

Comparação Tamanho dos Clados												
Com TBE												
	Distância	Clados	Tamanho (nº de clados → nº de indivíduos)									
CP_C	2.5	194	173→2	19→3	2→4							
	4.5	407	302→2	81→3	15→4	5→5	2→6	1→7	1→10			
	6.0	503	288→2	118→3	45→4	30→5	10→6	7→7	2→8	1→9	1→10	1→11
CP_B	2.5	55	44→2	10→3	1→4							
	4.5	93	69→2	15→3	7→4	1→5	1→6					
	6.0	129	85→2	21→3	15→4	6→5	1→6	1→10				
CP_F1	2.5	15	15→2									
	4.5	28	24→2	3→3	1→5							
	6.0	35	25→2	7→3	1→4	1→5	1→6					
CP_CRF31	2.5	12	9→2	1→3	1→4	1→5						
	4.5	27	18→2	5→3	3→4	1→5						
	6.0	26	14→2	5→3	2→4	2→5	2→7	1→8				
Sem TBE												
	Distância	Clados	Tamanho (nº de clados → nº de indivíduos)									
CP_C	2.5	195	174→2	19→3	2→4							
	4.5	441	325→2	90→3	17→4	5→5	2→6	1→7	1→10			
	6.0	576	312→2	148→3	55→4	34→5	14→6	6→7	2→8	1→9	3→10	1→11
CP_B	2.5	56	44→2	11→3	1→4							
	4.5	99	75→2	15→3	7→4	1→5	1→6					

	6.0	141	91→2	25→3	17→4	6→5	1→6	1→10				
CP_F1	2.5	15	15→2									
	4.5	28	24→2	3→3	1→5							
	6.0	37	27→2	7→3	1→4	1→5	1→6					
CP_CRF31	2.5	12	9→2	1→3	1→4	1→5						
	4.5	35	22→2	7→3	4→4	1→5	1→6					
	6.0	30	8→2	8→3	4→4	3→5	3→6	3→7	1→8			

Suplementar 2

Tabela 11: Comparação do tamanho dos *clusters* obtidos para os subtipos B, C, F1 e CRF31_BC a partir de árvores com e sem TBE.

Identificação de <i>clusters</i> de transmissão com Cluster Picker			
	Distância	Sem TBE	Com TBE
CP_C	2.5	195	194
	4.5	441	407
	6.0	576	503
CP_B	2.5	56	55
	4.5	99	93
	6.0	141	129
CP_F1	2.5	15	15
	4.5	28	28
	6.0	37	35

	2.5	12	12
CP_CRF31	4.5	35	27
	6.0	30	26
